





Rapid Training Data Generation for Tissue Segmentation Using Global Approximate Block-Matching with Self-organizing Maps

Lee B. Reid^(✉)  and Alex M. Pagnozzi 

The Australian e-Health Research Centre,
Commonwealth Scientific and Industrial Research Organisation,
Brisbane, QLD, Australia
lee.reid@csiro.au

Abstract. Deep learning techniques for tissue segmentation require large amounts of data for training, testing, and cross-validation. Manually generating such segmentations, however, is extremely time-consuming. This can lead to such techniques being limited to imaging modalities and populations for which ground truths already exist, overfitting, or the use of data that is not expert-checked and so likely to contain errors. A need exists for a means of accelerating expert tissue-segmentation, such as automated techniques that require little correction and have little reliance on existing atlases. Here, we describe a method which can reliably perform registration-free tissue-segmentation using a single atlas that is only partially complete. This Global Approximate Block-matching method utilizes a self-organizing map, an unorthodox artificial neural network. This network trains quickly only on the provided partial atlas and allows these labels to be propagated throughout the target image via block-matching. Using this technique we segmented brains of 22 subjects and compared its performance to expert ground truths. When provided with an atlas for which only 2% of voxels were labelled, this achieved mean dice similarity coefficients of 0.88 (grey-matter) and 0.92 (white matter). Performance improved as higher amounts of atlas were provided, up to a maximum of 0.93 (grey-matter) and 0.96 (white matter) when a single whole-brain atlas was provided. Although segmentations produced by this technique are sufficiently accurate to be used directly for many purposes, its primary use case may lie in accelerating the creation of expert atlases for deep-learning techniques.

Keywords: Block-matching · Self-organizing map · Segmentation
Ground-truth

1 Introduction

Deep learning techniques are gaining popularity for tissue segmentation but require large amounts of data for training, testing, and cross-validation.

© Crown 2018

D. Stoyanov et al. (Eds.): DLMIA 2018/ML-CDS 2018, LNCS 11045, pp. 110–118, 2018.

https://doi.org/10.1007/978-3-030-00889-5_13

Generating data ultimately requires manually delineated segmentations, but this process can take several days to complete per volume if attention to detail is desired. Although data augmentation can artificially boost the quantity of training data, this is unlikely to produce a satisfying dataset for cross-validation and does not necessarily provide the true anatomical variance seen in the population. As a result, segmentation techniques such as deep-learning are at risk of becoming limited to certain magnetic resonance (MR) sequences and populations for which ground truth data already exist. A need exists for a means of accelerating expert tissue-segmentation. One option is to automatically generate segmentations with a method that has little-or-no requirement on existing atlases, and to correct this segmentation as needed. A widely available tool to achieve this is expectation-maximization (EM) segmentation, but its accuracy and ability to self-improve when provided with human-corrected data can be limited.

Block-matching (BM) techniques are typically designed to remove image noise but can also perform tissue segmentation [2]. Such methods typically match similar cubes of voxels (patches) from an atlas to a target image, and compute a ‘non-local mean’ of these patches. Alternatively, labels from these patches can be averaged (rather than voxel intensities) to generate a probabilistic tissue segmentation. Block matching leverages the small patterns that exist throughout an image, such as the repeated sulcal folding of brain tissue. However, such techniques are unable to take full advantage of this redundancy because exhaustively comparing patches to one another is computationally expensive. To circumvent this, techniques such as volBrain [2] rely on atlas-to-target registration and search only a local area for patches similar to a target. This requires at least one whole-brain atlas with reasonable spatial correspondence to the target image.

Dimensionality reduction is an alternative, or complementary, way to reduce the computational cost of comparing patches exhaustively. The self-organizing map (SOM) is a neural-network based non-linear dimensionality reduction technique [1]. Briefly, SOMs are implemented as a collection of nodes which each have local connectivity, a fixed position in low dimensional space (e.g. forming a 2D grid), and a trainable position in the high-dimensional space. Through competitive learning, rather than backpropagation, SOMs train quickly and provide a smooth projection between high- and low-dimensional spaces. Intuitively, a trained 1D SOM can be thought to optimally ‘snake’, much as principal components analysis (PCA) draws a straight line, through high dimensional space.

We have developed a ‘Global Approximate Block-Matching’ (GAB) denoising and segmentation algorithm. GAB requires no spatial correspondence between the atlas and target, nor for the atlas to be completely labelled. This allows a partially segmented image to act as an atlas for another image, or to act as its own atlas, propagating manually segmented labels to non-segmented regions. To achieve this, GAB performs a whole-image search for atlas patches matching each target. To reduce this operation’s computational burden, each patch is collapsed into a singular value (SV) through a method such as the SOM. Here, we describe GAB and demonstrate its tissue segmentation performance using incomplete

atlases. The accuracy and speed afforded by this method may enable rapid atlas building, in turn enabling deep learning methods to target diverse populations and utilize MR sequences for which training data are not yet available.

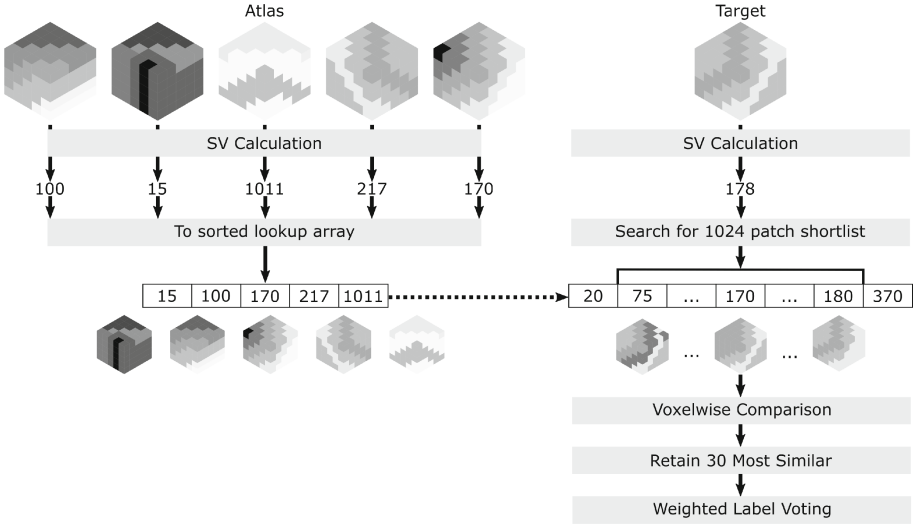


Fig. 1. In Step 1 (top left), the atlas image was split into overlapping $5 \times 5 \times 5$ voxel patches. For each patch, a singular value (SV) was calculated using one of four methods. Patches and their corresponding labels (not shown) were then sorted by these SVs. In Step 2 (right), for each target patch from a target image, an SV was calculated. Using a binary search, 1024 atlas patches with similar SVs were selected as a 'shortlist'. The voxel-wise sum of square differences (SSDs) were calculated for these patches versus the target, and the 30 patches with the most similar SSDs selected, their labels contributing toward final image reconstruction. See the text for details on final image reconstruction.

2 Methods

We tested the ability of GAB and EM to perform a series of three-tissue (cortical grey-matter, cortical white-matter, cerebrospinal fluid) segmentations in MR images. Our dataset consisted of N4 bias-corrected MPRAGE images (0.9 mm isotropic) from 23 participants ($28.8 \pm 1.5y$) acquired in a previous study [4, 5]. MR acquisition was approved by the local ethics committee. Participants gave written informed consent. We also utilized the expert (manually corrected) brain masks and expert tissue segmentations for each image generated during this study. GAB does not require that all areas of an atlas have accompanying labels; its segmentation accuracy was tested when provided with an atlas in varying degrees of completion (Fig. 3). Performance was judged by the quantitative similarity between automatically generated and expert generated segmentations.

2.1 The “Global Approximate Block Matching” Method

Images were processed in two steps: independently-applied denoising of both target and atlas images, followed by segmentation of the target image. Both steps used the GAB method. Below we detail how segmentation was performed, followed by a brief explanation as to how this was modified to perform denoising.

The GAB method, summarized in Fig. 1, accepts five images: (1) a target image, such as a T1; (2) a target mask; (3) an atlas image; (4) an atlas mask; and (5) atlas labels. The masked target is linearly intensity scaled to match the histogram of the masked atlas and both are stored as 8-bit unsigned integer images. These images are split into overlapping $5 \times 5 \times 5$ voxel patches within their respective brain masks. For each patch, an SV is calculated from voxel intensities. Atlas patches are then sorted by their SV. To find matching patches to a target, GAB conducts a binary search for the most similar atlas patch, based on target and atlas patch SVs. This approximate best-match, and those patches between 512 positions before and 511 positions after it in the sorted array, constitute a 1024-patch ‘shortlist’ of items likely to be similar to the target. The voxelwise sum of square differences (SSD) was calculated between the target patch and shortlist to identify the 30 most similar patches to the target. The labels for these patches are multiplied by their patch’s weight ($1/(SSD + 10^{-6})$), filtered by a Gaussian of $\sigma = 1$ voxel, and added to the appropriate label’s ‘sum’ image. These weights are multiplied by this Gaussian and added to a ‘weights’ image. Upon completion of all block matching, each sum image is divided by the weights image to generate a final tissue probability map. This ‘unweighting’ is required as each sum image voxel is contributed to by up to 125 block-matching operations, each operation in turn summing 30 weighted patches. A voxelwise maximum likelihood approach converts the probabilistic tissue maps into a hard segmentation.

2.2 Singular Value Calculation

Four GAB variants were tested, differing from one another by their SV calculation method: PCA (ϵ_0), mean voxel intensity, random (SV randomly generated), and SOM. Each SOM was arranged as 4096 nodes equally-spaced in 1D, and trained on up to 10^7 randomly selected patches from the input image. SV calculation using the SOM was performed by locating a patch’s continuous position in this array (i.e. between the best matched node and its most similar neighbor) based on voxel-wise SSD. PCA transformations were calculated from the same randomly selected patches. We also re-ran GAB-SOM with an artificially boosted number of atlas patches, providing the algorithm with 48 unique augmentations (all rotations, plus their mirror images) of each labelled atlas patch.

Denoising utilized GAB with two modifications to the method detailed above: (1) patches were $3 \times 3 \times 3$ in size; (2) the target, atlas, and label images were the same, I.E. the method matched patches within the target image to others within that same image. As such, it reconstructed a single low-noise version of the input image, rather than several probabilistic tissue maps. Denoising was

always performed with the SOM SV method, the performance of which was not quantified, as it is beyond the intended scope of this paper.

2.3 Expectation Maximization

For a comparator method, we used an Expectation Maximization segmentation algorithm with a modified Markov Random Field implementation. This method was selected as it has previously been reported to perform robustly in the absence of atlas based priors [3]. EM was executed with a single Gaussian per tissue class, initialized with means of 0, 2, and 3 for cerebrospinal fluid (CSF), grey-matter (GM), and white-matter (WM) respectively, each with $\sigma = 1$. These values were selected after empirical testing demonstrated that they produced reliable segmentation performance in a similar dataset acquired on the same scanner. Moderate deviations from this initialization did not meaningfully alter the performance of EM for the current dataset.

2.4 Atlas and Performance Metric

We use the term ‘atlas availability’ herein to refer to the fraction of an atlas’ labels which were made available to the segmentation algorithm. One randomly selected image was assigned as an atlas; the remaining 22 images constituted targets for segmentation. This atlas was converted into a series of ‘partially complete’ atlases, which were then used by GAB to segment targets. This was performed as follows: (1) $11 \times 11 \times 11$ voxel masks were placed on the atlas in the left temporal lobe, right temporal lobe, and frontal lobe, constituting the atlas labels mask (Fig. 3); (2) for each target image, the whole brain was segmented using only the atlas labels within this masked area and the result saved; (3) the labels mask was dilated with 6 connectivity and cropped to the brain mask. Steps 2 and 3 were repeated until the labels mask was identical to the brain mask, providing segmentations for each image across a range (0.2%–100%) of atlas availabilities. Dice similarity coefficients (DSC) for cortical GM and WM were calculated, within the entire brain mask, for each target segmentation by comparison to that target’s corresponding expert segmentation.

3 Results

Methods were implemented in .Net 4.0 and OpenCL 1.2 and ran on a Dual Xeon 8-core E5-2650 node with 128 GB of RAM and 3 Kepler Tesla K20 GPUs. Denoising + segmentation with GAB took 7–11.5 min in total, with processing using more-complete atlases taking longer than with incomplete atlases. EM segmentation ran in <1 min in each case. EM segmentation achieved DSCs of 0.67 ± 0.21 (*mean* \pm *SD*; GM) and 0.84 ± 0.19 (WM). All GAB methods except GAB-Random outperformed EM segmentation at atlas availability $\geq 0.8\%$. This accuracy increased markedly until 6% atlas availability, after which

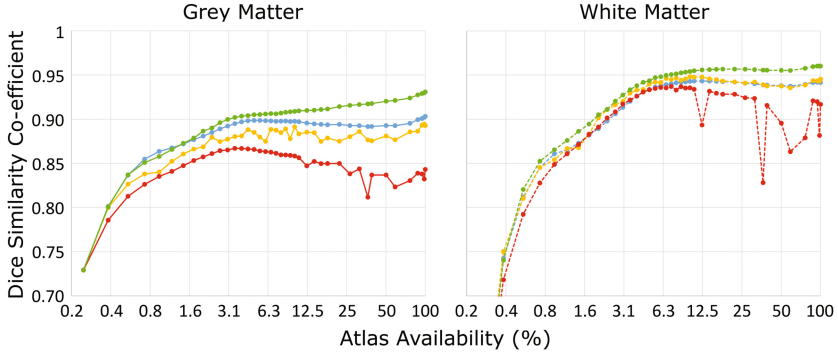


Fig. 2. Dice similarity coefficients for grey (left) and white (right) matter for segmentations generated by GAB, when provided with differing proportions of atlas. GAB methods are color-coded by their SV method as follows: Red, Random; Gold, PCA; Blue, Mean; Green, SOM. All methods achieved a dice similarity coefficient of 0.51 for white matter segmentation at an atlas availability of 0.24%, not shown here. (Color figure online)

Table 1. Dice similarity coefficients for GAB-derived grey matter (GM) and white matter (WM) segmentations at four different atlas availabilities. Each row indicates a different singular-value (SV) calculation method. SOM-48 indicate SOM-based SV calculation, with 48 patch augmentations (see text). All standard deviations were <0.02 , except GAB-Random which demonstrated SDs of 0.03 (GM) and 0.02 (WM) at 100% atlas availability.

SV	GM				WM			
	0.9%	2%	6%	100%	0.9%	2%	6%	100%
SOM-48	0.88	0.90	N/A	N/A	0.92	0.88	N/A	N/A
SOM	0.84	0.89	0.91	0.93	0.91	0.85	0.95	0.96
Mean	0.86	0.88	0.90	0.90	0.89	0.86	0.94	0.94
PCA	0.84	0.87	0.87	0.89	0.90	0.85	0.94	0.95
Random	0.84	0.86	0.86	0.84	0.89	0.85	0.94	0.91

a gradual increase was seen (Fig. 2; Table 1). GAB-SOM provided superior segmentation accuracy to other methods, particularly for GM labelling, with more stable results than GAB-PCA or GAB-Random (Fig. 2). When the SOM-based analyses were run with 48 augmentations of each atlas patch, the atlas availability required to achieve a DSC of 0.90 in both tissue classes fell from 3.1% to 1.7%. Such augmentation, however, was infeasible in the current implementation above 3.5% atlas availability because of GPU memory constraints.

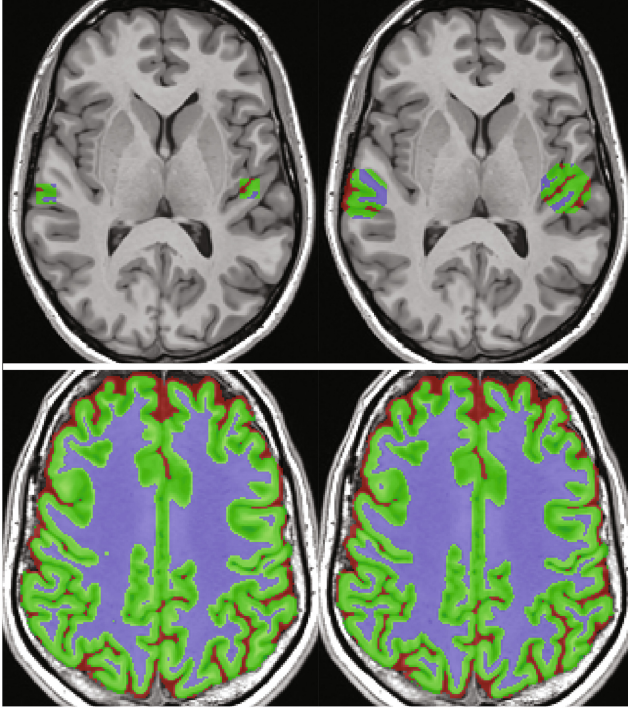


Fig. 3. Top: The atlas cropped to the labels-mask at 0.2% (left) and 2% (right) availability. The third labelled region is not visible in this slice. Bottom: Segmentation results for a representative dataset. The left segmentation was generated using GAB-SOM with 48 augmentations at 2% atlas availability. The right segmentation was generated with GAB-SOM at 100% atlas availability.

4 Discussion

Artificial neural networks such as deep learning can require large amounts of data for training, validation, and cross-validation in order to demonstrate task proficiency. In the case of brain-tissue segmentation, this often means that a large number of whole-brain tissue segmentations are required, but the time cost of generating these accurately can be very high. Here, we have demonstrated a Global Approximate Block-matching method which, unlike most methods, can segment a full brain MR image with reasonable accuracy when provided with an atlas that is predominantly incomplete. We found that this method reliably outperformed EM, an alternative technique with similar advantages, when provided with an atlas for which $\geq 0.8\%$ of voxels had been manually labelled. GAB was most effective when using an SOM for SV calculation, achieving dice coefficients of ≥ 0.9 for both cortical GM and cortical WM when provided with an

atlas that was as low as 1.7% complete (Fig. 2). GAB-SOM also demonstrated performance comparable with some deep learning networks when provided with a whole brain atlas [6]. The relative performance of GAB-SOM is likely due to the SOM's highly non-linear nature enabling an effective whole-brain search for similar patches to a target. This is indicated by the relatively poorer performance of GAB when relying on PCA or mean voxel intensity for SV calculation, particularly at moderate atlas availabilities.

One advantage of GAB, for generation of 'ground truth' segmentations, is that it can be used in an iterative strategy in which an image is automatically segmented, then partially manually corrected, in a repeated manner. In such a strategy, a target image would act as its own atlas, and the GAB-based segmentation can be expected to improve with each iteration. This has the potential to drastically lower the time-cost of generating the first 'ground truth' segmentation of a series. For segmentations of subsequent images, GAB is likely to perform a high-quality segmentation, as this first image can be provided as an atlas.

A block-matching segmentation algorithm, volBrain, has previously been described [2]. Presently, volBrain and GAB have different strengths. Whilst volBrain relies on multiple whole-brain atlases in order to perform multi-atlas label fusion, GAB requires only a fraction of an atlas to be provided. This makes GAB a stronger candidate for creating expert segmentations for new populations and imaging modalities. GAB also does not limit patch searches to a local area. This means it is not reliant on image registration, and may perform sensibly when target and atlas anatomy differ meaningfully, such as with pathology. However, unlike volBrain, modifications are likely to be needed for accurate delineation of localized tissues such as the deep grey matter. Potential modifications exist, such as including a patch's location as parameters in the SV calculation, or splitting volumes into regions which are segmented using different partial atlases, but these modifications are yet untested.

In conclusion, we proposed a Global Approximate Block-matching method that relies on the SOM as a powerful dimensionality reduction technique. When provided with minimal training data, this method generates accurate brain tissue segmentations that have little need for manual correction. This technique may prove a useful tool for quickly generating training data sets for deep learning methods targeting imaging modalities and populations for which ground truth data are not widely available.

References

1. Kohonen, T.: Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**(1), 59–69 (1982)
2. Manjon, J.V., Coupe, P.: volBrain: an online MRI brain volumetry system. *Front. Neuroinformatics* **10**, 30 (2016)
3. Pagnozzi, A.M., et al.: Alterations in regional shape on ipsilateral and contralateral cortex contrast in children with unilateral cerebral palsy and are predictive of multiple outcomes. *Hum. Brain Mapp.* **36**03, 3588–3603 (2016)

4. Sale, M., Reid, L., Cocchi, L., Pagnozzi, A., Rose, S., Mattingley, J.: Brain changes following four weeks of unimanual motor training: evidence from behavior, neural stimulation, cortical thickness, and functional MRI. *Hum. Brain Mapp.* **38**(9), 4773–4787 (2017)
5. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**(6), 1310–20 (2010)
6. Wachinger, C., Reuter, M., Klein, T.: DeepNAT: deep convolutional neural network for segmenting neuroanatomy. *NeuroImage* **170**, 434–445 (2018)