UNSUPERVISED REGION-BASED ANOMALY DETECTION IN BRAIN MRI WITH ADVERSARIAL IMAGE INPAINTING

Bao Nguyen*, *Adam Feldman*‡, *Sarath Bethapudi*†, *Andrew Jennings*†, *Chris G. Willcocks**

*Department of Computer Science, Durham University
‡Department of Engineering, University of Exeter
†County Durham and Darlington NHS Foundation Trust

ABSTRACT

Medical segmentation is performed to determine the bounds of regions of interest (ROI) prior to surgery. By allowing the study of growth, structure, and behaviour of the ROI in the planning phase, critical information can be obtained, increasing the likelihood of a successful operation. Usually, segmentations are performed manually or via machine learning methods trained on manual annotations. In contrast, this paper proposes a fully automatic, unsupervised inpainting-based brain tumour segmentation system for T1-weighted MRI. First, a deep convolutional neural network (DCNN) is trained to reconstruct missing healthy brain regions. Then, upon application, anomalous regions are determined by identifying areas of highest reconstruction loss. Finally, superpixel segmentation is performed to segment those regions. We show the proposed system is able to segment various sized and abstract tumours and achieves a mean and standard deviation Dice score of 0.771 and 0.176, respectively.

Index Terms— Anomaly Detection, Unsupervised Deep Learning, Generative Adversarial Networks, Inpainting

1. INTRODUCTION

Brain tumours are a growth of abnormal cells in the brain. Cancerous primary tumours are the most likely to cause severe symptoms, as they spread and cause damage to the surrounding normal brain tissue, raising intracranial pressure. Symptoms range from headaches and nausea to seizures, paralysis, and loss of vision and speech. Additionally, these tumours may cause a hemorrhagic stroke and can result in death within two days. Hence, early discovery is critical to prevent severe degeneration and allows for more treatment options, increasing the chance of recovery.

Segmentation is an important prerequisite to surgery. Manual segmentation requires radiologists to first determine the ROI, then manually to draw the boundaries surrounding the ROI. Although this is possible in 3D, this is usually done in a slice-by-slice manner by encircling the ROI or annotating voxels of interest. Not only is this time consuming but also prone to subjectivity and results are subject to large intra and inter expert variability, leading to considereable differences in extracted radiomic values [1]. There is a need for automatic methods to reduce cost, time and bias.

Contribution. This paper proposes a region-based anomaly detection method using image inpainting to generate a coarse heatmap, followed by a refinement postprocessing stage. Specifically, we adversarially train a DCNN architecture to locate and inpaint missing healthy brain regions. The intuition is that the network will fail to reconstruct unhealthy data that it has not observed in training. We then apply superpixel segmentation and select the segment with the highest reconstruction loss (based on the heatmap), resulting in a automated system that can better capture irregular tumours.

Related Work. Many unsupervised anomaly detection methods operate by training on a set of normal samples only. Anomalies are then determined if they lie outside the manifold of the learned representation. AnoGAN [2] was first to introduce Generative Adversarial Networks (GANs) [3] for anomaly detection which was then applied to retinal fluid segmentation in optical coherence tomography images. AnoGAN [2] introduced a new mapping scheme from high-dimensional normal images to the latent space representation (inverse-mapping). Anomalous regions can then be identified by computing the pixel difference between both images which inevitably induces false positives. Manual thresholding is then often required to reduce them. Another drawback of AnoGAN [2] is that the mapping procedure is an expensive iterative process, defined as the minimization through $\gamma = 1, 2, \ldots, \Gamma$ backpropagation steps, thus requiring Γ optmization steps, which result in poor inference-time performance. Following AnoGAN [2], other works [4, 5, 6] have improved upon instance anomaly detection and the computationally expensive mapping procedure. However, these works do not offer the ability to localize anomalous regions.

Inpainting for anomaly detection has been less extensively investigated. Works from [7, 8, 9] have been successfully applied to pedestrian segmentation, on-road anomaly segmentation and surface inspection domains. These results indicate success in applications where the domain is consistent and predictable, e.g. with many patterns with small variability.

2. METHODOLOGY

Brain tumour segmentation is performed in three steps (see Fig. 1): i) Adversarially train a DCNN to reconstruct missing healthy brain regions. ii) Given a query slice, perform a masked sliding window operation to obtain predictions for all regions. iii) Construct a heatmap indicating areas of highest reconstruction loss and perform superpixel segmentation to better capture object boundary.

The inpainting network consists of two concurrently trained sub-networks, a generator G trained to capture the data distribution of normal brain MRI and a discriminator D trained to classify real slices from G's reconstructions. Both G and D play the minimax two-player game, such that G aims to maximize the probability of D making a mistake. In effect, D acts as a critic, ensuring G produces realistic reconstructions. The training process is as follows: i) Given a random normal 2D brain MRI slice, x, a square mask of length γ is erased at a random location with the mask function $\Psi(x,\gamma)$. ii) The result of $\Psi(x,\gamma)$ is then passed to G for reconstruction, $G : \Psi(x,\gamma) \to \hat{x}$. Where $x, \Psi(x,\gamma), \hat{x} \in \mathbb{R}^{w \times h \times c}$. iii) G's weights are then updated from a combined reconstruction and adversarial loss. To learn the contextual information of the normal training data, L1 loss is applied to x and G's reconstruction.

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x \sim p_x} \left[||x - G(\Psi(x,\gamma))||_1 \right]. \tag{1}$$

For G to produce realistic reconstructions the following adversarial loss is incorporated to the overall learning objective:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{x \sim p_x} [\log(1 - D(G(\Psi(x,\gamma))))]. \tag{2}$$

Thus giving the complete learning objective for G:

$$\mathcal{L}_{\text{G}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}, \tag{3}$$

where λ_{rec} and λ_{adv} are weighting parameters denoting the importance of reconstruction and adversarial loss respectively. iv) Finally, D's weights are updated with the following:

$$\mathcal{L}_{\text{D}} = \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{x \sim p_x} [\log(1 - D(G(\Psi(x,\gamma))))]. \tag{4}$$

To update the network parameters via the Adam optimizer algorithm, the "update" function is employed.

Architecture Details. Inspired by UNet [10] and in contrast to [11], G inherits skip-connections such that each down-sampling layer in the contracting path is concatenated to its corresponding up-sampling layer in the expansive path. These skip-connections allow for better quality reconstructions, increasing convergence speed and stabilizing training. In high-dimensional spaces, the density ratio estimation by the discriminator is often inaccurate and unstable during training, causing the generator to fail in learning the multimodal structure of the target distribution. Thus, for the discriminator we

---

Algorithm 1: Train Inpainting Network

1   $\boldsymbol{\theta}_{\text{G}}, \boldsymbol{\theta}_{\text{D}}$ = initialise network parameters
2   $\lambda_{\text{rec}}, \lambda_{\text{adv}}$ = initialise weighting parameters
3   $\alpha_{\text{G}}, \alpha_{\text{D}}$ = initialise network learning rates
4   γ = initialise mask size
5   **repeat**
6     $\mathbf{x}$ = random mini-batch from dataset
7     $\hat{\mathbf{x}} = G(\Psi(\mathbf{x},\gamma))$
8     $\mathcal{L}_{\text{rec}} = ||\mathbf{x} - \hat{\mathbf{x}}||_1$
9     $\mathcal{L}_{\text{adv}} = \log(1 - D(\hat{\mathbf{x}}))$
10    $\boldsymbol{\theta}_{\text{G}} = \text{update}(-\nabla_{\boldsymbol{\theta}_{\text{G}}} \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}, \alpha_{\text{G}})$
11    $\boldsymbol{\theta}_{\text{D}} = \text{update}(\nabla_{\boldsymbol{\theta}_{\text{D}}} \log(D(\mathbf{x})) + \log(1 - D(\hat{\mathbf{x}})), \alpha_{\text{D}})$
12   **until** *converged*

---

apply Spectral Normalization [12], rescaling the weights with spectral norm σ by the power iteration method.

Localization System. To obtain predictions for all regions in x, the localization system employs the sliding window algorithm $\phi(x,\gamma,k)$, with a window size (square mask) of length γ and a window step size of k pixels. Each frame produced in $\phi(x,\gamma,k)$ is passed to the inpainting network for reconstruction. To reduce the impact of image contrast on reconstruction loss [13], we apply minmax normalization $w(x)$ on x and $G(\phi(x,\gamma,k))$, such that $w(x) : [\min(x), \max(x)]^{W \times H \times D} \mapsto [0, 255]^{W \times H \times D}$, where $\min(x)$ and $\max(x)$ are the minimum and maximum pixel values of x respectively. After the normalization process, we compute the reconstruction (L1) loss between the normalized original window and normalized reconstructed window only. The result is a heatmap indicating regions that exhibit the most reconstruction loss.

To automate the segmentation process, we apply superpixel segmentation on the query slice x and select the segment with the most reconstruction loss, based on the constructed heatmap. Superpixel based techniques segment an image into a collection of connected pixels which are similar in colour and texture. This allows us to capture common features in our slice and locate features our inpainting network failed to reconstruct. Although there exists many algorithms that perform superpixel segmentation, we choose to implement Felzenszwalb's efficient graph-based segmentation [14] as it segments tumours more accurately than other methods.

Dataset. The Neurofeedback Skull-stripped (NFBS) [15] repository provides 125 manually skull-stripped T1-weighted MRI scans of normal brain tissue only, simplifying the inpainting task. Due to tumours being easier to identify, 2D input slices of size 256×192 are of the axial view of the brain. For evaluation, we utilise a tumorous dataset of 22 T1-weighted MRI scans provided by the Centre for Clinical Brain Sciences from the University of Edinburgh [16]. The dataset consists of non-skull-stripped scans of 2D axial dimensions 256×156.
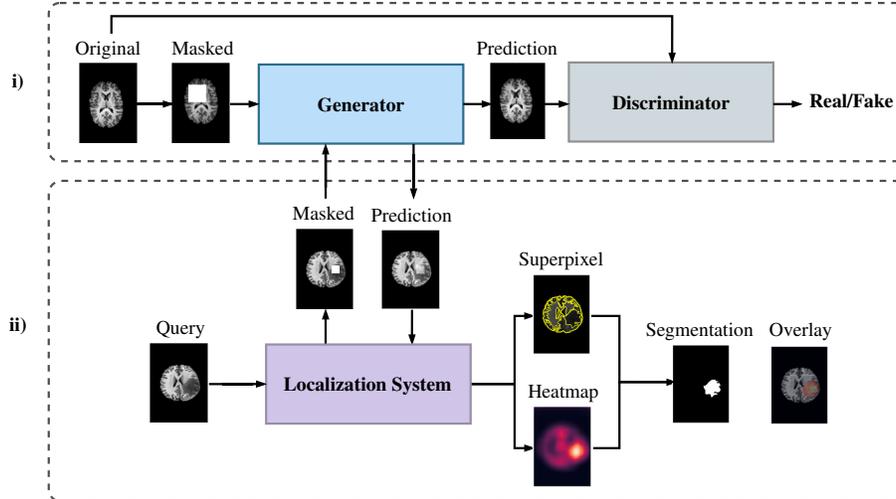
Fig. 1: The Proposed System - i) Train Inpainting Network. ii) Predict anomalous regions and obtain segmentation.

Preprocessing. In addition to standardization and normalization for the tumour dataset, we utilise BrainSuite [17, 18, 19] for skull-stripping and bias field correction to rectify image intensity non-uniformities that are a result of magnetic field variations rather than anatomical differences. Furthermore, we add padding (black pixels) to each slice in the tumour dataset so that it matches the same dimensions as the normal dataset. Note, in every scan in the normal training set, the brain is already surrounded by black pixels, so not foreign to the inpainting network.

3. RESULTS

Evaluation was performed on γ values of 8, 16, 32 and 64, with a k step value of 4 pixels on separately trained networks. To evaluate the inpainting performance and segmentation accuracy, we extracted 100 slices from a normal test set [15] and 100 slices from the tumour dataset [16] (where tumour is visible), respectively. The inpainting network was trained for 30,000 iterations in minibatches of size 16; learning rates for G and D were set to 0.0001 and 0.001; $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{adv}}$ were set to 50 and 1 respectively, as these settings empirically showed consistent optimal performance. We compared the segmentation performance with AnoGAN [2], where AnoGAN [2] was trained for 60,000 iterations in minibatches of size 8.

As shown in figure 2, one can see how the problem definition changed according to the amount of data masked. For smaller window sizes, the inpainting network is only required to reconstruct a handful of pixels. However, with larger masks the network is required to reconstruct texture and structure. Additionally, due to adversarial training, the network produces reconstructions without blur.

In figure 3, the heatmaps indicate higher reconstruction loss values where the tumour is present. Heatmaps produced from smaller window sizes performed better on healthy re-
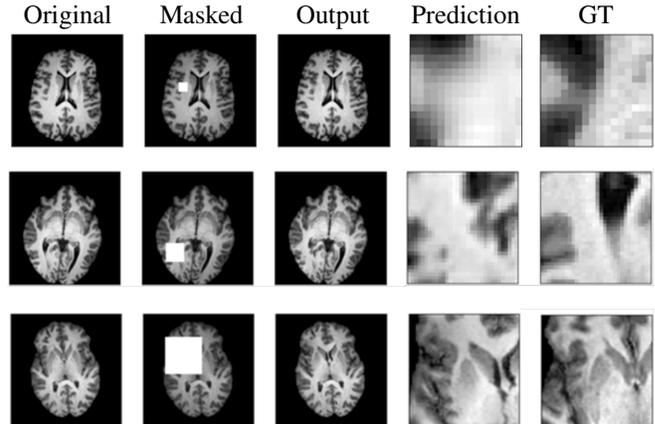

Fig. 2: Inpainting results for window sizes 16, 32 and 64 respectively. From top to bottom.

	PSNR ($\mu \pm \sigma$)	SSIM ($\mu \pm \sigma$)	Dice ($\mu \pm \sigma$)
Ours - 8	38.61 ± 3.159	0.99 ± 0.006	0.49 ± 0.396
Ours - 16	37.94 ± 4.664	0.98 ± 0.022	0.66 ± 0.346
Ours - 32	35.63 ± 5.413	0.97 ± 0.002	$\mathbf{0.77 \pm 0.176}$
Ours - 64	31.66 ± 6.491	0.96 ± 0.003	0.55 ± 0.380
AnoGAN [2]	-	-	0.38 ± 0.238

Table 1: Inpainting performance on unseen healthy samples and Dice scores.

gions, whereas larger window sizes appear to localize the tumour more effectively. It is important that the window covers most if not all the tumour to distinguish the reconstruction performance with other healthy areas of the brain. However, if the size of the tumour is known beforehand, then choosing the "best-fit" window would be preferable.
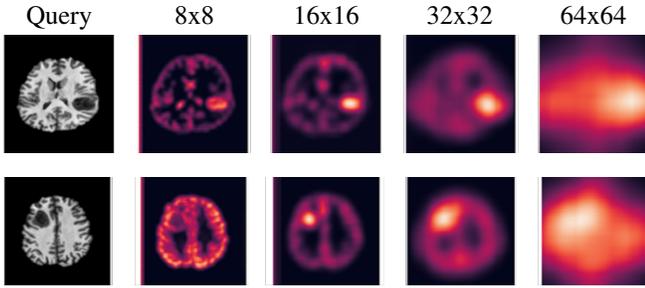
Fig. 3: Heatmap for various window sizes.
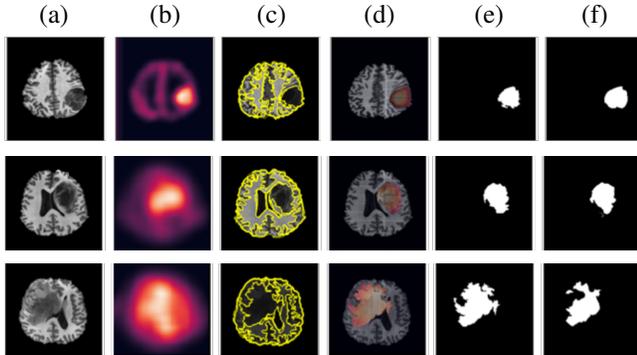
Fig. 4: Segmentation Results - (a) Query, (b) Heatmap, (c) Superpixel, (d) Overlay, (e) Prediction, (f) Ground Truth

Segmentation Quality. As shown in table 1, the proposed system outperforms AnoGAN [2]. A window of size 32 indicates the highest accuracy and consistency in handling various sized tumours, with a mean Dice score of 0.771 and of standard deviation 0.176. Other window sizes express high variability, performing well either on small or large sized tumours. The effect of using superpixel results in higher definition segmentations, allowing the capture of regular and more abstract tumours, which would not be possible with manual thresholding. Additionally, for evaluation, we trained a single network to reconstruct all values of γ (8, 16, 32 and 64). We obtained good reconstruction results, but this resulted in partial reconstructions of tumorous regions during the sliding window process, indicating a single optimal network for all tumour sizes is not possible with our method.

4. DISCUSSION

Figure 5 illustrates snapshots of the sliding window process. The inpainting network restores the masked region to what a normal brain would look like in that area. The inpainting network is naive and does not reconstruct the tumour when partially covered, due to computing L1 loss globally rather than just on the local masked region.

Localization is only possible if the inpainting network performs well on non-anomalous regions, but fails otherwise. Additionally, the overall quality of segmentations depends
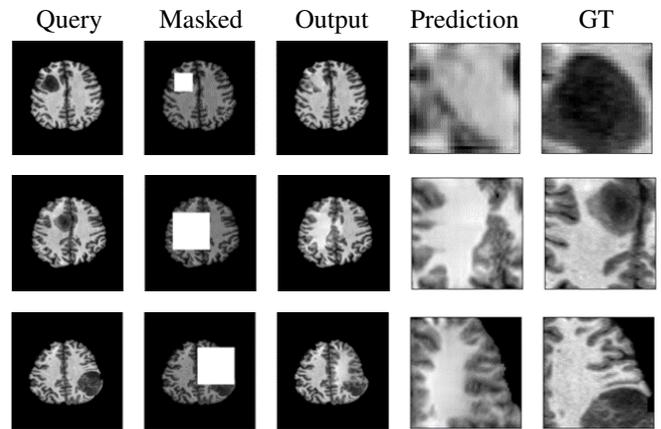
Fig. 5: Examples of full and partial tumour coverage.

on the superpixel algorithm. If tumours are not substantially different in colour and texture, no segmentation would be possible.

In the future, we would like to investigate extending both Felzenszwalb's algorithm and our inpainting approach to handle 3D volumes. This is challenging due to the memory usage and computational overhead associated with 3D volumes.
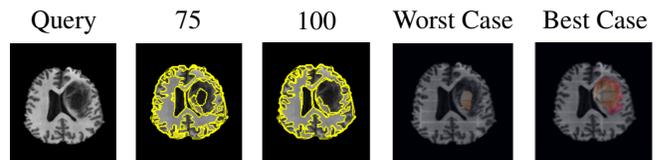
Fig. 6: Superpixel segmentation with different scale values and possible outcomes.

In Felzenszwalb's algorithm [14], the scale parameter influences the segment size. Although a constant scale value of 75 was used in the evaluation process, there were times when a larger or smaller value would have produced better segmentation results. Figure 6 illustrates how a scale value below 100 would result in the inner segment being selected. To prevent this, an appropriate scale value must be manually chosen before the segmentation process.

5. CONCLUSION

The proposed system for unsupervised anomaly detection in the application of T1-weighted brain MRI outperforms AnoGAN [2] for both the mean and standard deviation Dice scores, where the system accurately segments regular and abstract tumours of varying size. The main limitations are where normal segments surrounding the tumour are detected as false positives. However, the approach of an initial coarse unsupervised region-based segmentation strategy, followed by a separate refinement stage, has proved to be effective in the application of unsupervised medical segmentation.

6. COMPLIANCE WITH ETHICAL STANDARDS

The training data used [15] in this research study was conducted retrospectively using human subject data made open access by the Nathan Kline Institute. The data was acquired in compliance with the Health Insurance Portability Act (HIPPA). The evaluation data used [16] was also made open access by the Centre for Clinical Brain Sciences from the University of Edinburgh.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Chintan Parmar, E. Rios Velazquez, R. Leijenaar, M. Jermoumi, S. Carvalho, R. Mak, S. Mitra, B. U. Shankar, R. Kikinis, Benjamin Haibe-Kains, P. Lambin, and H. Aerts, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLoS ONE*, vol. 9, 2014.

[2] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *IPMI*, 2017.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[4] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar, "Efficient gan-based anomaly detection," *ArXiv*, vol. abs/1802.06222, 2018.

[5] Samet Akçay, Amir Atapour Abarghouei, and Toby P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," *ArXiv*, vol. abs/1805.06725, 2018.

[6] Samet Akçay, Amir Atapour Abarghouei, and Toby P. Breckon, "Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection," *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2019.

[7] Ji Hoon Joung, Michael S. Ryoo, Sunglok Choi, and Sung-Rak Kim, "Reliable object detection and segmentation using inpainting," *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3871–3876, 2012.

[8] Asim Munawar and Clement Creusot, "Structural inpainting of road patches for anomaly detection," *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 41–44, 2015.

[9] Matthias Haselmann, Dieter P. Gruber, and Paul Tabatabai, "Anomaly detection using deep learning based image completion," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1237–1242, 2018.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," *ArXiv*, vol. abs/1505.04597, 2015.

[11] Deepak Pathak, Philipp Krähenbühl, J. Donahue, Trevor Darrell, and Alexei A. Efros, "Context encoders: Feature learning by inpainting," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, 2016.

[12] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida, "Spectral normalization for generative adversarial networks," *ArXiv*, vol. abs/1802.05957, 2018.

[13] Amanda Berg, Jorgen Ahlberg, and Michael Felsberg, "Unsupervised learning of anomaly detection from contaminated image data using simultaneous encoder training," *ArXiv*, vol. abs/1905.11034, 2019.

[14] Pedro F. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, 2004.

[15] Benjamin Puccio, James P. Pooley, John Pellman, Elyse C Taverna, and Richard Cameron Craddock, "The preprocessed connectomes project repository of manually corrected skull-stripped t1-weighted anatomical mri data," *bioRxiv*, 2016.

[16] Cyril Pernet, Krzysztof Gorgolewski, and Whittle Ian, "A neuroimaging dataset of brain tumour patients," 2016.

[17] D. Shattuck and R. Leahy, "Brainsuite: An automated cortical surface identification tool," *Medical Image Anal.*, vol. 6, pp. 129–142, 2002.

[18] S. Sandor and R. Leahy, "Surface-based labeling of cortical anatomy using a deformable atlas," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 41–54, 1997.

[19] D. Shattuck, Stephanie R. Sandor-Leahy, K. Schaper, D. Rottenberg, and R. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage*, vol. 13, pp. 856–876, 2001.