High Resolution, Unbiased CT Brain Template

John Muschelli[a]

[a]*Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, 615 N Wolfe St, Baltimore, MD, 21205*

Abstract

Clinical imaging relies heavily on X-ray computed tomography (CT) scans for diagnosis and prognosis. Many research applications aim to perform population-level analyses, which require images to be put in the same space, usually defined by a population average, also known as a template. We present an open-source, publicly available, high-resolution CT template. With this template, we provide voxel-wise standard deviation and median images, a basic segmentation of the cerebrospinal fluid spaces, including the ventricles, and a coarse whole brain labeling. This template can be used for spatial normalization of CT scans and research applications, including deep learning. The template was created using an anatomically-unbiased template creation procedure, but is still limited by the population it was derived from, an open CT data set without demographic information. The template and derived imaes are available at https://github.com/muschellij2/high_res_ct_template.

1. Introduction

Many research applications of neuroimaging use magnetic resonance imaging (MRI). MRI allows researchers to study a multitude of applications and diseases, including studying healthy volunteers as it poses minimal risk. Clinical imaging, however, relies heavily on X-ray computed tomography (CT) scans for diagnosis and prognosis. Studies using CT scans cannot generally recruit healthy volunteers or large non-clinical populations due to the radiation exposure and lack of substantial benefit. As such, much of head CT data is gathered from prospective clinical trials or retrospective studies based on health medical record data and hospital picture archiving and communication system (PACS). Most of this research is on patients with neuropathology, which can cause deformations of the brain, such as mass effects, lesion, stroke, or tumors.

Many clinical protocols perform axial scanning with a high within-plane resolution (e.g.0.5mm x 0.5mm) but lower out-of-plane resolution (e.g. 5mm). High resolution scans (out of plane resolution ≈ 0.5mm) may not be collected or reconstructed as the lower resolution scans are typically those read by the clinician or radiologist for diagnosis and prognosis. Recently, a resource of a large number of CT scans were made avaiable, denoted as CQ500 (Chilamkurthy

---

et al. 2018). These scans include people with a number of pathologies, including hemorrhagic stroke and midline shifts. Fortunately, this data also includes people **without indicated pathology** with high resolution scanning, which is what we will use in this study.

The goal of this work is to create an anatomically unbiased, high-resolution CT template of the brain. That is, we wish to create a template that represents the population, regardless of any initial templates we start with. The first, and we believe the only, publicly-available CT template was released by Rorden et al. (2012) (https://www.nitrc.org/projects/clinicaltbx/). This template was created with the specific purpose of creating a template with a similar age range as those with stroke, using 30 individuals with a mean age of 65 years old (17 men). The associated toolbox released contained a high resolution (1x1x1mm) template, with the skull on, in Montreal Neurological Institute (MNI) space. Subsequent releases have included skull-stripped brain templates, but only in a lower (2x2x2mm) space (https://github.com/neurolabusc/Clinical).

Thus, the current CT template available are a high-resolution template ($1mm^3$), but not of the brain only (and skull stripping the template performs marginally well), and a low-resolution template of the brain only, both in MNI space. We have used these templates in previous analyses, but would like a brain template that was 1) constructed using an unbiased anatomical procedure, 2) uses more patients, 3) uses high-resolution scans to achieve a higher resolution, and 4) provide an image which dimensions are easily used in deep learning frameworks.

As the CQ500 data was released under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC-NC-SA) International License, we can release the template under the same license.

1.1. Methods

All code, analysis, and reporting was done the R statistical programming language (R Core Team 2015) and a number of packages from the R medical imaging package platform Neuroconductor (Muschelli et al. 2019).

1.1.1. Data

We defined a high-resolution patient scan as having a within-axial resolution of 0.7x0.7mm or less, with full coverage of the brain. For example, if the cerebellum was not imaged, that image was discared. All scans were non-contrast CT scans with a soft-tissue convolution kernel. As CT scans are generally well calibrated across sites and are measured in standard units of Hounsfield Units (HU), no intensity normalization was done. Intensities less than -1024 HU (the value for air) and greater than 3071 HU were Winsorized to those values, as values outside of these are likely artifact or areas outside the field of view.

All data was converted from DICOM files to NIfTI (Neuroimaging Informatics Technology Initiative) using `dcm2niix` (Li et al. 2016) using the `dcm2niir` package (Muschelli 2018). This conversion corrects for any gantry tilt and enforces one fixed voxel size for the image, which is necessary if different areas of

the image are provided at different resolutions, which is sparsely seen in clinical CT images.

From the CQ500 data set, 222 subjects had no indication of pathology, of which 141 had a high-resolution scan (if multiple were present, the one with the highest resolution was used). From these 141 people, 130 had "thick-slice" scans where the out-of-plane resolution was greater than 4mm. We used these 130 scans for construction of the template. The 11 scans were discarded as we wish to perform the same operation using low-resolutions scans to see the effect of initial resolution on template creation, but that is not the focus of this work.

We chose an image (patient 100 from CQ500) with a resolution of 0.488x0.488x0.5mm. The head was skull-stripped so that only brain tissue and CSF spaces were kept, using a previously validated method (Muschelli et al. 2015) using the brain extraction tool (BET) from FSL (FMRIB Software Library) (Smith 2002; Jenkinson et al. 2012). The image was resampled to 0.5x0.5x0.5mm resolution so that the voxels are isotropic. As most head CT have a within-axial dimension of 512x512 and we would like the image to be square, we padded the image back to 512x512 after resampling, and the image had 336 coronal-plane slices. This image was used for template creation.

1.1.2. Template Creation

The process of template creation can be thought of as a gradient descent algorithm to estimate the true template image as inspired by the advanced normalization tools (ANTs) software and the R package ANTsR that implements the registration and transformation was used (https://github.com/ANTsX/ANTsR). The process is as follows:

1. Let I_i represent the image where i represents subjects. We will Register all images to the template, denoted $\bar{T}_k$ where k represents iteration, using an affine registration followed by symmetric normalization (SyN), a non-linear deformation, where the composed transformation is denoted as $G_{i,k}$ (Avants et al. 2008). Let the transformed image be denoted as $T_{i,k}$. In other words, $I_i \overset{G_{i,k}}{\to} T_{i,k}$. The transformation $G_{i,k}$ is represented by a 4D warping image. Let T_1 be the original template chosen above and $G_{i,1}$ be the transformation for an image to the original template.

2. Calculate a the mean, median, and standard deviation images, where the mean image is $\bar{T}_k = \frac{1}{n} \sum_{i=1}^{n} T_{i,k}$, using a voxel-wise average.

3. Calculate the average warping transformation: $\bar{G}_k = \frac{1}{n} \sum_{i=1}^{n} G_{i,k}$. A gradient descent step size of 0.2 was specified for SyN gradient descent, such that: $\bar{T}_{k+1} = \bar{T}_k \times \left(-0.2 * \bar{G}_k \right)$. The median and standard deviation are transformed accordingly.

For each iteration k, we can calculate a number of measures to determine if the template has converged compared to the previous iteration $k - 1$. We calculated the Dice Similarity Coefficient (DSC) between the mask of iteration k

and $k-1$, where the mask for iteration k is defined as $\bar{T}_k > 0$. We also the root Mean mquared error (RMSE) of voxel intensities. The RMSE can be calculated over a series of areas, either 1) the entire image, 2) over the non-zero voxels in iteration k, 3) in iteration $k-1$, or 4) the union (or intersection) of the 2 masks. Calculation over the entire image gives an optimistic estimate as most of the image are zeroes, and the choice of either iteration k or $k-1$ masks is arbitrary, so we calculated the RMSE over the union of the 2 masks.

To define convergence, we would like a high DSC between the masks and a low RMSE. Ideally, the convergence criteria would set a DSC of 1 and a RMSE less than 1 Hounsfield Unit (HU), which would indicate the voxel intensity is changing less than 1 HU on average. As CT scans are measured in integers, this RMSE would likely be as good as possible. We set a DSC cutoff of 0.95 and chose the template with the lowest RMSE. As this procedure is computationally expensive, we ran 40 iterations, which was adequate for achieving stable results (Figure 1).

Values of the final template that were lower than 5 HU were boundary regions, outside the region of the brain and likely due to singlular segmentation images, incongruent with the remainder of the image (Supplemental Figure 4). We did not contrain the DSC and RMSE calculation excluding these regions, but excluded values less than 5 HU from the final template.

After the template was created, we padded the coronal plane so that the template was 512x512x512. The intention is that these dimensions allow it easier to create sub-sampled arrays that are cubes and multiples of 8, such as 256x256x256 or 64x64x64 with isotropic resolution.

1.1.3. Segmentation

Though the template itself is the main goal of the work, many times researchers use or are interested in annotations or segmentations of the template space. In many CT scans, the contrast between gray matter and white matter is not high, except for specific structures such as the cerebellum, corpus callosum, and basal ganglia. Thus, segmentation methods based on intensity may not differentiate gray and white matter adequately. We used a multi-atlas registration approach. We used a previously published set of 35 MRI atlases from Landman et al. (2012), which had whole brain segmentations, including tissue-class segmentations.

We registered each brain MRI with the template using SyN and applied the transformation to the associated tissue segmentation and whole brain segmentation from that template. Thus, we had 35 tissue segmentations of the CT template in template space, and the segmentations were combined using STAPLE (Warfield, Zou, and Wells 2004) via the `stapler` package (Muschelli 2019). The whole brain structures were combined using majority vote.

In addition, we segmented the template using Atropos (Avants et al. 2011) Separating the brain from the cerebrospinal fluid areas (mainly ventricles) are of interest in many applications, such as Alzheimer's disease [Leon et al. (1989); braak1999neuropathology].
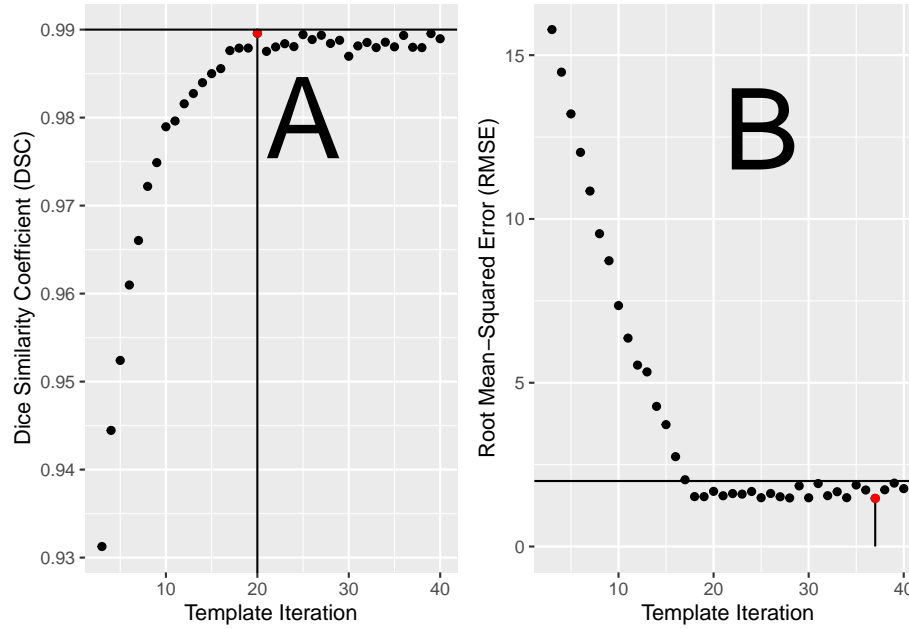
4

Figure 1: Convergence of Shape and Intensity of the Template over Iterations. Here we see the Dice Similarity Coefficient (DSC) increase between an iteration and the previous iteration, achieving high degrees of overlap, indicating the shape of the surface of the image is similar and converging (panel A). We also see the root mean-squared error (pane) drops as the iterations increase and then levels off around 4 Hounsfield units (HU), the horizontal line. The red dot indicates the iteration chosen for the template.

1.2. Results

As we see in Figure 1A, the DSC quickly increases and reaches a high score, where the horizontal line indicates a DSC of 0.99. The red dot and vertical line indicate the iteration that had the maximum DSC (0.9896). As the DSC is high for all iterations past iteration 15, we chose the template based on the minimum RSE. In Figure 1B, we see a similar pattern of improving performance, but by lowering the RMSE. The lowest RMSE is noted by the red point with a value of 1.47. Thus, this iteration (iteration 37) is the template we will choose.

The template for this image can be seen in Figure 2, along with the standard deviation image, and a histogram of the intensities of the template. We see the template is relatively smooth, with values from 5 HU to around 65 HU. The standard deviation image shows high variability around the lateral horns, which may be due to calcifications in a set of patients, which have abnormally high HU values.

```
## Warning in text.default(x = 65, y = 1e+05, text = "C", cex = 2): "text" is not a
## graphical parameter
```
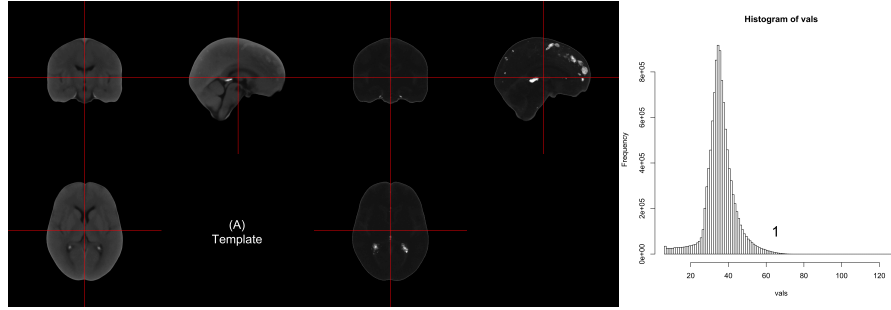
See Figure**??**

5

Figure 2: Template Image, Standard Deviation Image, and Histogram of Intensities. Here we show the template in the left panel, the voxel-wise standard deviation, denoting areas of variability (which include biological and technical variability), and the histogram of the template intensities/Hounsfield Units (HU). Overall the template is smooth and values fall in the range of 5 to 65 HU.
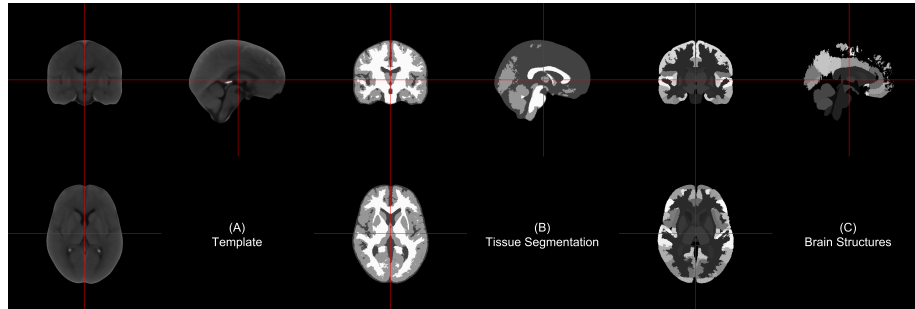

Figure 3: Template Image, Tissue Segmentation, and Whole Brain Segmentation. We see the areas of white matter, gray matter, cebebrospinal fluid (CSF) in Panel B. We see

1.3. Discussion

We present a high-resolution, publicly-available CT template with associated segmentations and other annotations of the template. The data used was from a publicly-available dataset, the CQ500. The main downside with the CQ500 data set is that no demographic or clinical information was released for each patient, save for indication for pathology. Therefore, we cannot attest the general population of interest for this template. Furthermore, we cannot fully assume these patients were disease-free as a lack of pathology only applies to the categories of interest in the CQ500 dataset (intracranial/subdural/subarachnoic/epidural hematoma, Calvarial or other fractures, mass effect and midline shifts). In future work, we hope to prepare age- and sex-specific templates for each population based on hospital scans and records, where we have demographic information and confirmation of lack of neuropathology.

In addition to the template, we have provided a set of segmentations. This includes a whole brain segmentation of over 150 structures. Though this may prove useful, we caution users to how well this template can provide an accurate segmentation of these structures. At least, the accuracy of the segmentation may have variable accuracy at different areas of the brain.

The resulting image dimensions was 512x512x512, with a resolution of 0.5x0.5x0.5mm. The fact that the image dimension is a multiple of 8 allows it to be resampled to 1x1x1mm and 2x2x2mm and remain as a cube. These dimensions are particularly important in certain deep learning architectures and frameworks. Though most templates are given using the mean image, we believe the standard deviation image represents variability in the area. This variability represents true systematic and biologic variability. One important area of systemic variability is registration errors. Therefore this template allows for the creation of z-score images, where a new image is registered to the mean image, the mean image is subtracted, and then divided by the standard-deviation image, so that voxels represent standard deviations away from the mean voxel. This image may be a useful tool in feature extraction. Thus, we believe this template provides a standard, isotropic space that is conducive to machine learning and can reduce the burden of standardization for medical imaging applications.

CQ500 is Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Therefore, the template is released under the same license.

References

2. Supplemental Material

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain." *Medical Image*
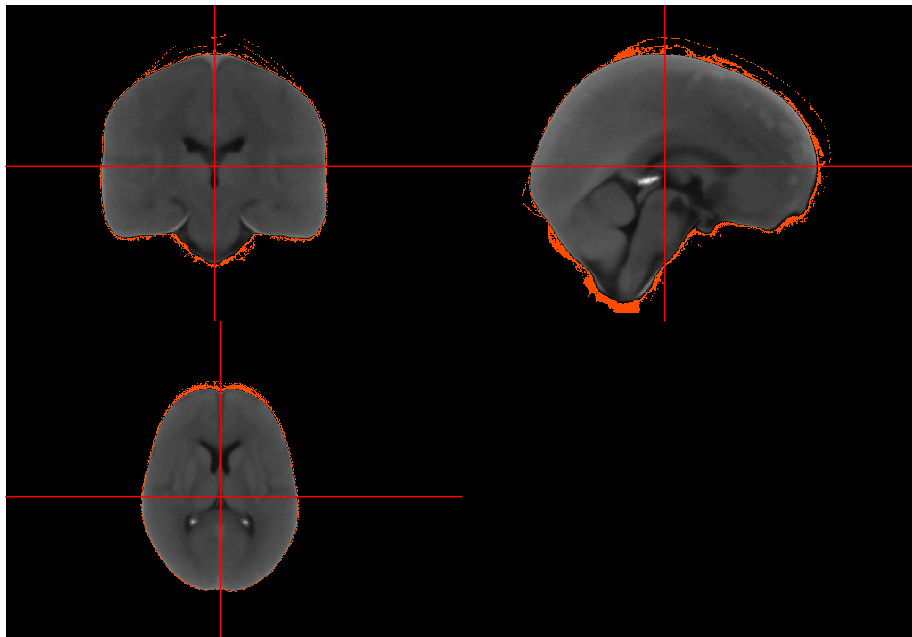
Figure 4: Boundary Issues with Low HU Values. Here we present the average image with the mask of voxels that were lower than 5 HU in the template. We excluded these values from the final template.

Analysis, Special issue on the third international workshop on biomedical image registration – WBIR 2006, 12 (1): 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Avants, Brian B, Nicholas J Tustison, Jue Wu, Philip A Cook, and James C Gee. 2011. "An Open Source Multivariate Framework for N-Tissue Segmentation with Evaluation on Public Data." *Neuroinformatics* 9 (4): 381–400.

Chilamkurthy, Sasank, Rohit Ghosh, Swetha Tanamala, Mustafa Biviji, Norbert G Campeau, Vasantha Kumar Venugopal, Vidur Mahajan, Pooja Rao, and Prashant Warier. 2018. "Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans: A Retrospective Study." *The Lancet* 392 (10162): 2388–96.

Jenkinson, Mark, Christian F. Beckmann, Timothy E. J. Behrens, Mark W. Woolrich, and Stephen M. Smith. 2012. "FSL." *NeuroImage* 62 (2): 782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.

Landman, Bennett Allan, Annemie Ribbens, Blake Lucas, Christos Davatzikos, Brian Avants, Christian Ledig, Da Ma, et al. 2012. *MICCAI 2012 Workshop on Multi-Atlas Labeling.* CreateSpace Independent Publishing Platform.

Leon, Mony J de, Ajax E George, Barry Reisberg, Steven H Ferris, Alan Kluger, Leonidas A Stylopoulos, Jeffrey D Miller, Mary Ellen La Regina, Clara Chen, and Jacob Cohen. 1989. "Alzheimer's Disease: Longitudinal CT Studies of Ventricular Change." *American Journal of Roentgenology* 152 (6): 1257–62.

Li, Xiangrui, Paul S Morgan, John Ashburner, Jolinda Smith, and Christopher Rorden. 2016. "The First Step for Neuroimaging Data Analysis: DICOM to NIfTI Conversion." *Journal of Neuroscience Methods* 264: 47–56.

Muschelli, John. 2018. *dcm2niir: Conversion of DICOM to NIfTI Imaging Files Through R.* https://www.nitrc.org/plugins/mwiki/index.php/dcm2nii:MainPage.

———. 2019. *Simultaneous Truth and Performance Level Estimation.* https://CRAN.R-project.org/package=stapler.

Muschelli, John, Adrian Gherman, Jean-Philippe Fortin, Brian Avants, Brandon Whitcher, Jonathan D Clayden, Brian S Caffo, and Ciprian M Crainiceanu. 2019. "Neuroconductor: An R Platform for Medical Imaging Analysis." *Biostatistics* 20 (2): 218–39.

Muschelli, John, Natalie L. Ullman, W. Andrew Mould, Paul Vespa, Daniel F. Hanley, and Ciprian M. Crainiceanu. 2015. "Validated Automatic Brain Extraction of Head CT Images." *NeuroImage.* https://doi.org/10.1016/j.neuroimage.2015.03.074.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rorden, Christopher, Leonardo Bonilha, Julius Fridriksson, Benjamin Bender, and Hans-Otto Karnath. 2012. "Age-Specific CT and MRI Templates for Spatial Normalization." *NeuroImage* 61 (4): 957–65. https://doi.org/10.1016/j.neuroimage.2012.03.020.

Smith, Stephen M. 2002. "Fast Robust Automated Brain Extraction." *Human Brain Mapping* 17 (3): 143–55. https://doi.org/10.1002/hbm.10062.

Warfield, Simon K, Kelly H Zou, and William M Wells. 2004. "Simultaneous Truth and Performance Level Estimation (Staple): An Algorithm for the Validation of Image Segmentation." *IEEE Transactions on Medical Imaging* 23 (7): 903–21.