Development of the next-generation functional neuro-cognitive imaging protocol - Part 1: A 3D sliding-window convolutional neural net for automated brain parcellation

Heath M. Lorzel[*], Mark D. Allen

Cognitive FX, 280 West River Drive, Suite 110, Provo, UT 84604, United States

A B S T R A C T

Functional MRI has emerged as a powerful tool to assess the severity of Post-concussion syndrome (PCS) and to provide guidance for neuro-cognitive therapists during treatment. The next-generation functional neuro-cognitive imaging protocol (fNCI2) has been developed to provide this assessment. This paper covers the first step in the analysis process, the development of a rapidly re-trainable, machine-learning, brain parcellation tool. The use of a sufficiently deep U-Net architecture encompassing a small ($39 \times 39 \times 39$ voxel input, $27 \times 27 \times 27$ voxel output) sliding window to sample the entirety of the 3D image allows for the prediction of the entire image using only a single trained network. A large number of training, validating, and testing windows are thus generated from the 101 manually-labeled Mindboggle images, and full-image prediction is provided via a voxel-vote method using overlapping windows. Our method produces parcellated images that are highly consistent with standard atlas-based methods in under 3 min on a modern GPU, and the single network architecture allows for rapid retraining (<36 hr) as needed.

1. Introduction

Mild traumatic brain injury (mTBI) represents approximately 70 % to 90 % of traumatic brain injuries in the United States with an incidence of 600 in 100,000 people per year (Cassidy et al., 2004, Faul et al., 2010). Though the terms mTBI and concussion are often used interchangeably, concussion represents a variety of mTBI characterized by the absence of structural brain damage, though clinical manifestations may be similar (McCrory et al., 2013). Concussive symptoms typically resolve in 7 to 10 days (sports-related concussions) or within 3 months (non-athletes) (McCrea et al., 2003). However, approximately 33 % of patients will have persistence of symptoms with 30 % of those patients meeting post-concussion syndrome (PCS) criteria 6 months out from time of injury (Mittenberg et al., 2001; Rimel et al., 1981; Binder et al., 1997).

Standard structural clinical neuroimaging studies demonstrate no abnormal findings for the majority of PCS patients as the clinical presentations of PCS are thought to be caused by cerebrovascular dysregulation and neuronal dysfunction (Leddy et al., 2007; Giza et al., 2001; Ellis et al., 2015). However, functional MRI (fMRI), which uses blood oxygen-level dependent (BOLD) signaling, has shown abnormalities in

patients with PCS (Giza et al., 2014; Eierud et al., 2014).

1.1. Functional neuro-cognitive imaging (fNCI)

Previous work (Allen et al., 2018) describes the development and use of a functional neuro-cognitive imaging (fNCI) protocol, based on the theory that dysregulated neurovascular coupling plays a dominant role in PCS, can be quantified through analysis of an abnormal BOLD signal pattern (Epps et al., 2017), and is used to identify biomarkers that serve to target interventions in PCS patients.

This protocol, retroactively referred to as fNCI1, consists of a set of six neuro-cognitive exercises that are derivations of conventional neuropsychological tests that have been redesigned for use inside of an MRI. Each of the tests employs a task-related design with several "active" phases where the subject is performing a cognitive task alternated with several "rest" phases where the subject is asked to count silently until the next active phase.

In the fNCI1 pipeline, functional images were acquired during each test, were spatially smoothed with a FWHM Gaussian kernel, and were co-registered to a high-resolution T1-weighted structural image. A time-series analysis of covariance in SPM12 was used to test each voxel

against the null-hypothesis that changes in the BOLD signal in that voxel do not significantly correlate with the temporal sequencing of the active-rest cycle of the cognitive test. A neuro-anatomical expert defined the boundaries of several functional regions on the structural image, and the voxel with the strongest functional response in each region was marked as the "activation peak".

The distribution of activation peaks in all regions was assessed for normality among a set of control subjects, and used to assemble a three-dimensional activation standard, or normative atlas, which was then used as a benchmark to quantify and localize the individual PCS patient activation patterns. These patterns were then organized into a set of five discovered biomarkers and a total combined PCS Biomarker score (Fig. 1).

1.2. Function neuro-cognitive imaging - Next Generation (fNCI2)

fNCI1 was initially made available for clinical use through Notus Neurological Imaging in 2008 and has been in use at Cognitive FX since 2014 to guide the treatment of several hundred PCS, Traumatic Brain Injury (TBI) and Acquired Brain Injury (ABI) patients. The next generation of fNCI intends to expand upon the original through the use of modern processing power, machine learning and statistical techniques, and advances in the understanding of functional neuro-anatomy. Specifically, the fNCI2 protocol intends to:

- Decrease the required time to obtain a scan by combining the six tests into three.
- Replace the need for a qualified neuro-anatomical expert to hand select functional region boundaries for each subject by developing an automated parcellation tool that can extract the regions of interest from a high-resolution 3D structural image that is co-registered to each individual's functional image.

- Improve the concept of the functional region 'activation peak' by using modern computational speed and machine-learning algorithms to analyze both individual and groups of voxels in each region.
- Greatly expand the number of regions examined per test (from about 6–10 to about 50).
- Expand the analysis of the cognitive tests to include network analysis on top of the functional region response.
- Organize the above into congruent 'Functional Systems', each of which performs a small specific task (eg. Target Search, Processing Speed, etc....) to facilitate their use by neuro-cognitive therapists.
- Completely automate the entirety of the above, with a reasonable computational run time, to allow for clinical use.

The remainder of this paper is devoted to the first step of this process; the development of the machine-learning brain parcellation tool, 'AutoParc'.

1.3. Automated parcellation

Accurate parcellation of cortical brain regions and segmentation of subcortical brain structures is a crucial first step to the fNCI pipeline, as well as for the future study of various brain disorders including schizophrenia, Parkinson disease, and multiple-sclerosis. For the purposes of the fNCI2 tool, 'AutoParc' must satisfy the following:

- Obtain a full parcellation of the required cortical regions and subcortical structures.
- Obtain a sufficiently accurate solution, one that will allow its use in obtaining the appropriate coarser-grid fMRI voxels, in a reasonable computational time on reasonable processing power such that it can be used in a clinical setting (ie 'a few minutes').
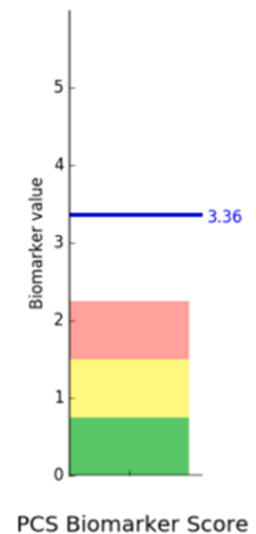- Accommodate individual variability in the cortical landmarks.

Fig. 1. A brief summary of the fNCI1 report data, showing the activation peaks for each of the regions analyzed in each of the six neuro-cognitive tests (left), along with the summarized biomarker scores (middle) and total PCS score summary (right).

- Allow for rapid re-training on new labeled data as knowledge of cortical structures and activation patterns evolve through use at the clinic.
- Extend the parcellation from the grey matter into the appropriate portions of adjacent white matter in order to include vascular structures that feed the cortical surface of a region, which is of critical importance for fMRI analysis.
- Utilize a loss-function and/or other corrective measures during training to bias the solution towards the elimination of false positives among detected regions. (The coarse grid of the overlain fMRI is more forgiving to false negatives).

A common approach to parcellation and segmentation is an atlas-based method, such as that available in the Freesurfer pipeline (Fischl, 2012). Atlas-based methods align one or several anatomical templates to the target image and then transfer segmentation and parcellation labeling from the template to the image. Due to the complexities of the registration step(s), these methods typically suffer from prohibitively long computational run times, often struggle to capture complex local anatomical variabilities, and can fail in the presence of large deformations.

Cortical-surface-based parcellation methods (Pantaxis et al., 2010) address local anatomical variabilities by treating the cortex as a convoluted surface and co-registering across subjects so that cortical landmarks and features are aligned, however these methods typically still require the intervention of a skilled neuro-anatomical expert, and/or continue to suffer from long computational run times due to the registration steps.

Supervised Deep Learning, most often in the form of Convolutional Neural Nets (CNNs) (LeCun et al., 1998), has become popular for both general, medical, and neural image segmentation. Unlike Atlas or Surface methods that require the use of hand-crafted features, deep learning techniques learn hierarchical features by training on large sets of manually labeled data, and can often be parallelized to run on modern GPUs to obtain reasonable computational times. Neural Nets are typically composed of multiple layers of several types, such as convolutional and pooling, with parameters that are learned through back-propagation. Convolutional Neural Nets have the further advantage of having layers that are connected to only a small number of units in adjacent layers which correspond to spatially localized regions, reducing the number of parameters in the net and limiting the computational requirements.

U-Net structures (Çiçek et al., 2016), which consist of several convolutional layers serving as 'encoders' followed by several de-convolutional layers serving as 'decoders'' produce an output image of approximately the same size as the input, have demonstrated state-of-the-art performance on several general imaging tasks. While extension of this tool to 3D images is computationally straightforward, practical limitations have thus far prevented a 3D U-Net from learning a full-brain parcellation by directly training on high-resolution brain MRIs. These limitations include memory and computational speed limitations, the low-number of available manually-labeled training sets, and the high number of labels.

Several pieces of recent work attempt to navigate these limitations by using modifications of 2D methods. "ParcelCortex" (Thyreau et al., 2020), uses 2D CNNs to reproduce parcellations on cortical ribbons. "DBPN" (Zhang et al., 2019), uses a two-stage deep network, consisting of a 2D coarse parcellation U-Net and a refinement network, to reproduce parcellations on a 2D cortical surface. "QuickNAT" (Roy et al., 2019) consists of three 2D F-CNNs operating on coronal, axial and sagittal views followed by a view aggregation step to infer the final segmentation. "FastSurfer" (Henschel et al., 2020) improves upon QuickNAT with the introduction of competition with each block, by replacing concatenation with maxout operations, and by the inclusion of a wider image context within each 2D-F-CNN by passing in neighboring slices instead of a single slice in order to retain a portion of the 3D information.

A 3D CNN patch-based architecture (Dolz et al., 2018) consisting of several convolutional layers followed by several fully connected layers and a final classification layer showed good performance in classifying subcortical structures. The design used small kernels, allowing for a deeper architectures while maintaining reasonable computational times, and intermediate connections to maintain multi-scale information. The loss of spatial information from the patching prevented its usefulness from expanding to cortical regions.

DeepNAT (Wachinger et al., 2018) uses a similar design, several convolutional layers followed by several fully connected layers, to classify the center voxel in a patch along with its neighboring voxels by adding a 3D fully-connected conditional random field. While it is also demonstrated against subcortical structures only, the spatial context is retained in the patches by augmenting the inputs with coordinate information.

SLANT (Huo et al., 2019) solves the problem of computational size by employing multiple independent 3D Fully Connected Networks (FCN), with each network being responsible for a particular spatial location before being joined in a label fusion technique. Similarly, AssemblyNet (Coupé et al., 2020) uses overlapping independently trained U-Nets, with certain U-Nets being initialized by their trained neighbors via transfer learning. Both methods show good results for cortical regions and subcortical structures.

UnesT (Yu et al., 2023), uses a nested transformer, U-shaped model with a fast and simplified encoder design that enables local communication between adjacent patches using hierarchical aggregation, which preserves relative location information. The problem of a relatively small number of manually labeled sets is addressed by augmenting the data set with T1 scans labeled by existing multi-atlas segmentation models.

In this work, we propose a simpler approach that utilizes a single U-Net to learn the entirety of the brain in a patch-based, sliding window method. The U-Net is relatively small compared to the size of the entire image and those used for SLANT and AssemblyNet, allowing both for the creation of a much deeper network architecture with small kernels, a greater number of filters, and a higher training batch size to improve convergence. Since the same U-Net will model multiple spatial locations, spatial concerns are addressed by encoding coordinate data into the 4th dimension and by using an input that is larger than the output. The single network aids in the problem of having a low number of training samples, as the training images are broken up into tens of thousands of training windows. The full brain is assembled by evaluating the U-Net on hundreds of overlapping windows, with each voxel being labeled by a majority vote.

While the method described in this paper is tuned for application to the fNCI2, the overall approach is highly suitable for general purpose brain parcellation and segmentation.

2. Methods

2.1. Training data and pre-processing

2.1.1. Labeled training data

The Mindboggle-101 dataset (Klein et al., 2017) contains manually parcellated and labeled brain surfaces and volumes derived from MRIs of 101 healthy individuals. The data set consists of several groups (OASIS-TRT-20, NKI-TRT-20, HLN-12, Afterthought, MMRR-3T7T, MMRR-21, NKI-RS-22, and Twins-2) and uses two complementary labeling protocols; the 'Desikan-Killiany-Tourville' cortical labeling protocol (31 labels per hemisphere) and FreeSurfer's non-cortical-plus-white-matter labels. We use the provided nifti volumes for both the structural and labeled image in their original space for each group.

2.1.2. Label adjustment and combination

Because automated parcellation is to be used to extract relevant features from a coarse-grid, overlain fMRI image and will be used by neuro-cognitive therapists to guide treatment, several adjustments are applied to the labels supplied by Mindboggle.

The required adjustments fall into one of three types. First, regions shown in Table 1 maintain the same anatomical layout as in Mindboggle but have numbers re-ordered. Second the Mindboggle regions shown in Table 2 are combined to form new labeled regions. Finally, boundaries between certain regions are moved slightly. A frontal plane is extended at the anterior most point of the caudal anterior cingulate and used to divide the portion of the superior frontal gyrus that lies near the medial plane. The posterior portion is assigned to the caudal anterior cingulate and the anterior portion to the rostral anterior cingulate. Similarly, a frontal plane is extended at the anterior most point of the paracentral gyrus, dividing the posterior cingulate and assigning the anterior portion to the caudal anterior cingulate. While not required for the demonstration of the parcellation, these optional steps were included as a component of the larger study. A version of AutoParc trained without these custom boundary adjustments is included in the following analysis.

Table 1

Anatomical regions taken directly from Mindboggle with re-ordered label numbers.

Label number	Label name
0	Unknown
1	Left thalamus proper
2	Right thalamus proper
3	Left caudal anterior cingulate
4	Left caudal middle frontal
5	Left fusiform
6	Left inferior parietal
7	Left inferior temporal
8	Left lateral occipital
9	Left lateral orbitofrontal
10	Left lingual
11	Left medial orbitofrontal
12	Left middle temporal
13	Left paracentral
14	Left pars orbitalis
15	Left postcentral
16	Left posterior cingulate
17	Left precentral
18	Left precuneus
19	Left rostral anterior cingulate
20	Left rostral middle frontal
21	Left superior frontal
22	Left superior parietal
23	Left supramarginal
24	Left insula
25	Right caudal anterior cingulate
26	Right caudal middle frontal
27	Right fusiform
28	Right inferior parietal
29	Right inferior temporal
30	Right lateral occipital
31	Right lateral orbitofrontal
32	Right lingual
33	Right medial orbitofrontal
34	Right middle temporal
35	Right paracentral
36	Right pars orbitalis
37	Right postcentral
38	Right posterior cingulate
39	Right precentral
40	Right precuneus
41	Right rostral anterior cingulate
42	Right rostral middle frontal
43	Right superior frontal
44	Right superior parietal
45	Right supramarginal
46	Right insula

Table 2

New regions consisting of compositions of Mindboggle regions.

Label number	Constituent labels	New label name
47	Left superior temporal, left transverse temporal	Left superior temporal
48	Right superior temporal, right transverse temporal	Right superior temporal
49	Left pars opercularis, left pars triangularis	Left inferior frontal gyrus
50	Right pars opercularis, right pars triangularis	Right inferior frontal gyrus
51	Left cuneus, left pericalcarine	Left cuneus
52	Right cuneus, right pericalcarine	Right cuneus
53	Left entorhinal, left parahippocampal	Left parahippocampal
54	Right entorhinal, right parahippocampal	Right parahippocampal
55	Left caudate, left putamen, left pallidum	Left basal ganglia
56	Right caudate, right putamen, right pallidum	Right basal ganglia
57	Left hippocampus, left amygdala	Left hippocampus
58	Right hippocampus, right amygdala	Right hippocampus

2.1.3. Dilations

Because AutoParc is intended to be an optimal parcellation approach for fMRI studies, it is important that the output labeled region include not only the grey matter surface of a functional region, but also the white matter tissue approximately enclosed within the cortical surface manifold(s) of each region. This allows for detection of surface vessels as well as supply vessels (e.g., penetrating arterioles) which contribute to the overall functional hyperemia measured from a typical region using reasonable voxel sizes for whole-brain fMRI BOLD detection. Specifically, we perform a 'closing' iteration of magnitude 4 (ie dilating each region 4 times and then eroding each region 4 times to remove small internal holes) followed by 2 dilation iterations in the direction of the white matter and 1 dilation iteration in any direction that doesn't contain white matter or another labeled region (Fig. 2). A version of AutoParc trained without these dilations is included in the following analysis.

2.1.4. Rotation

Each group of the Mindboggle-101 scans has the brain set in a different orientation of the i-j-k axis. For consistency, a rotation matrix is applied to each of the scans and the affines such that i-j-k has a consistent alignment with the sagittal-frontal-coronal directions.

2.1.5. Bias field correction

The T1 weighted image of each scan was corrected for the low-frequency intensity non-uniformity using the SimpleITK N4 Bias Field Correction (Beare et al., 2018).

2.1.6. Brain extraction

The brain was stripped from the skull and surrounding tissue using the open source "DeepBrain" extractor tool (Iitzco, 2018) by creating a probability mask and applying a cutoff of 0.5. The extracted brain is used to obtain the eigenvectors for the orientation (next step), which were then superimposed back onto the full images to continue the processing.

2.1.7. Orientation and normalization

To allow for consistency across different MRI hardware, the structural image of the extracted brain is normalized using the mean and standard deviation of the intensity of the voxels to a mean of 0 and standard deviation of 1.

The affines were recomputed for consistency of size and alignment by first setting the origin to be at the center of mass, as determined from the intensity of the structural image of the extracted brain. Then, the inertial tensor of the extracted brain was calculated and used to obtain the eigenvectors (essentially using the intensity of each voxel as a
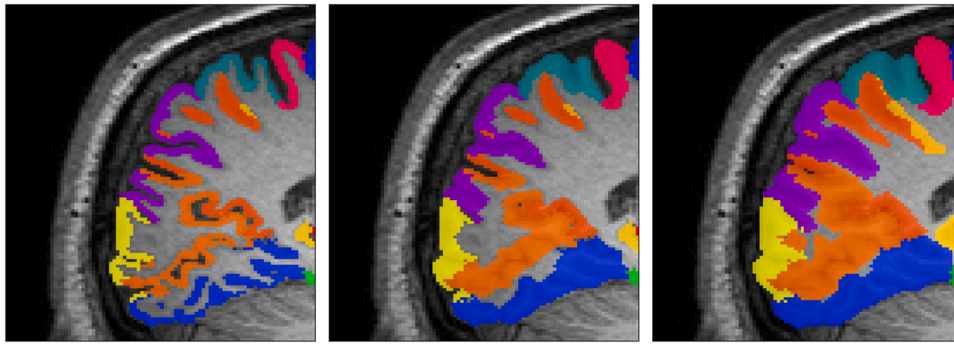
Fig. 2. A section of cortical labels shown overlain on the structural image as provided by Mindboggle (left); after performing four closing iterations (center); and after performing 4 closing iterations, two dilation iterations into the white matter, and one dilation iteration into the remainder of the unlabeled space (right).

discretized 'mass' for the inertial tensor calculation), which were then used to set the x, y, and z directions in the i, j, and k grid space. Finally, the size of the brain was non-dimensionalized to the width, which was determined as the furthest extent of the x-axis in both directions as measured from the center of mass.

2.1.8. Encoding of orientation data

Each of the training scans is to be broken down into a set of overlapping, sliding, 3D windows to generate the training images that will be fed into the neural net. In order to avoid the loss of spatial information, we encode the x,y,z data into the 4th dimension on the data. The method is similar to how color in two-dimensional images is represented by a 3rd dimension as [R,G,B] channels, but in our case we have a three-dimensional image with the 4th dimension as [Voxel-Intensity, X, Y, Z], with the latter three being produced from the new affine data generated in the orientation step.

2.1.9. Generation of training windows

Sliding windows of target size (27,27,27) voxels are swept across the training brains to generate the training images. Each window overlaps with the previous one by approximately 50 % in order to capture each voxel in approximately 8 different sliding windows. Further, the sliding window of the structural image is padded by 6 voxels on every side in order to capture information adjacent to the targeted voxel, giving a window size of (39, 39, 39) for the structural image sliding windows and (27, 27, 27) for the labeled image sliding windows.

2.1.10. Filtering

Given that a great deal of the generated sliding window images are likely to contain entirely or mostly 'unknown' labeled voxels, we scrub the majority of these in order to train the neural net to focus on the regions of interest. Specifically, we randomly scrub 95 % of the sliding window images that contain greater than 99.99 % 'unknown' voxels. We lose very little information this way, given that each voxel exists in approximately 8 different sliding windows. The entirety of the image pre-processing is shown in Fig. 3.

2.2. Artificial neural net structure

The goal of the artificial neural net structure is to learn the segmentation provided by a set of manually labeled (27,27,27) voxel windows from the set of T1 image (39,39,39,4) windows. To this end, we constructed a U-Net architecture of the type commonly used in medical image segmentation.

The U-Net is constructed using Google's Tensorflow and consists of a set of 18 'Convolutional Blocks' followed by 12 'Deconvolutional Blocks'. Each Convolutional Block consists of a 3D convolutional layer, a batch normalization layer, and an activation layer. Similarly, each Deconvolutional Block consists of a 3D transposed convolutional layer, a batch normalization layer, and an activation layer. Each block uses a kernel size of (3,3,3) and contains an increasing number of filters and decreasing window size as the layers reach deeper into the 'U'.

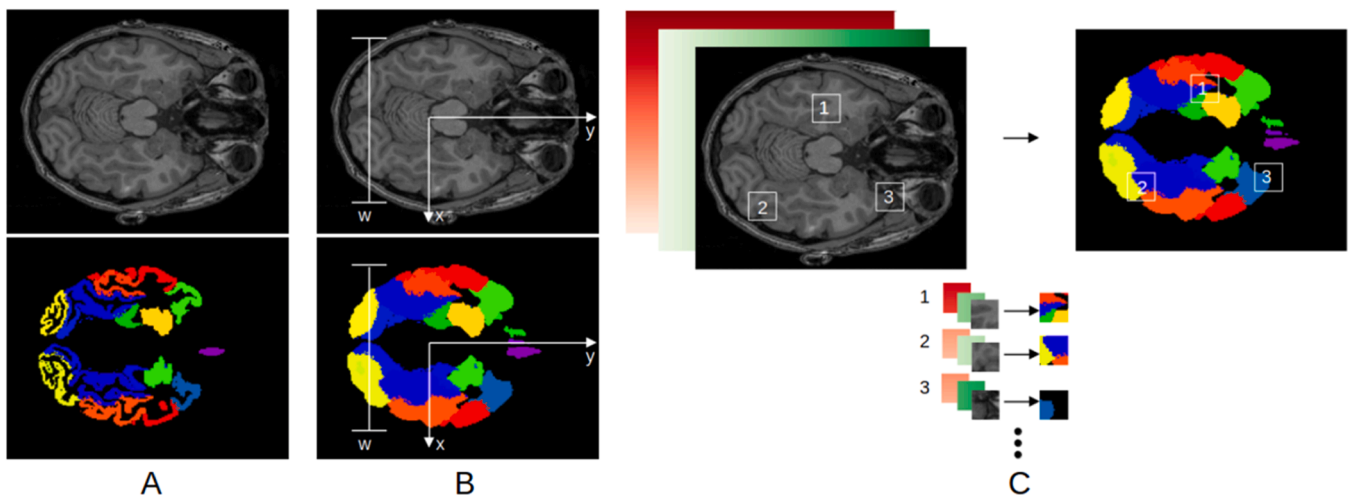An input layer of the same dimensions as the training images

Fig. 3. A summary illustration of the pre-processing steps. The MindBoggle labels are re-assembled into the desired combinations for training and aligned with their T1 structural images (A). The images are rough aligned for consistent i-j-k and then more finely aligned using the extracted brain eigenvectors, normalized to the width for consistent size, and the labels are dilated appropriately (B). Finally, the orientation data is encoded in the 4th dimension (illustrated here as a color sweep) and training windows are generated from the image (C). A 2D version of a 3D concept is shown for simplicity.

(39,39,39,4) followed by a batch normalization layer lies upstream of the first convolutional block. Each Convolutional Block accepts the outputs of the preceding block as its inputs (eg. Convolutional Block 12′s input is the output from Convolutional Block 11. Information flows down into the 'U') while each Deconvolutional Block accepts the outputs of both the preceding block and the Convolutional Block of equivalent dimensions as input (eg. Deconvolutional Block 12′s input is the sum of the output from Deconvolutional Block 13 and Convolutional Block 13. Information flows both up from the 'U' and across from the other side of the 'U'). In this manner, both large scale and fine scale information is preserved. The last Deconvolutional Block has one final layer of size (27,27,27) with a Softmax activation attached (Fig. 4).

Each training sample was fed into the U-Net in batches of size 24. The remaining hyperparameters consisted of an L2 regularization weight of 1.0e-6, an exponential linear unit activation function, and the Adam optimizer (Table 3).

A custom loss function, referred to as Combo Loss (CL), was designed consisting of a weighted combination of the sparse categorical cross entropy loss (SCCE) and the soft dice loss (SD) (Milletari et al., 2016)

$$Loss_{CL} = (1 - w) \cdot Loss_{SCCE} + w \cdot Loss_{SD}$$

SCCE typically has a faster convergence with a cleaner gradient, while the Dice Loss has a slower convergence, a more complicated gradient that is more likely to become unstable, but often obtains superior results. The combination of the two can result in improved performance during training, and the weighting of each has recently been tested in several papers. In this work, we use a custom weighting based on the current value of the SD during each training epoch,

$$w = \left(1 - Loss_{SD}^2\right)$$

The end result is an initial fast convergence from the dominance of the SCCE, followed by a slower but more thorough convergence as the SD takes dominance in the later training epochs with the smaller SCCE remaining in the equation to serve as a stabilizing term. The combination of the two into a single loss function resulted in a smoother training history, and an improved total accuracy, while maintaining reasonable overall computational times.
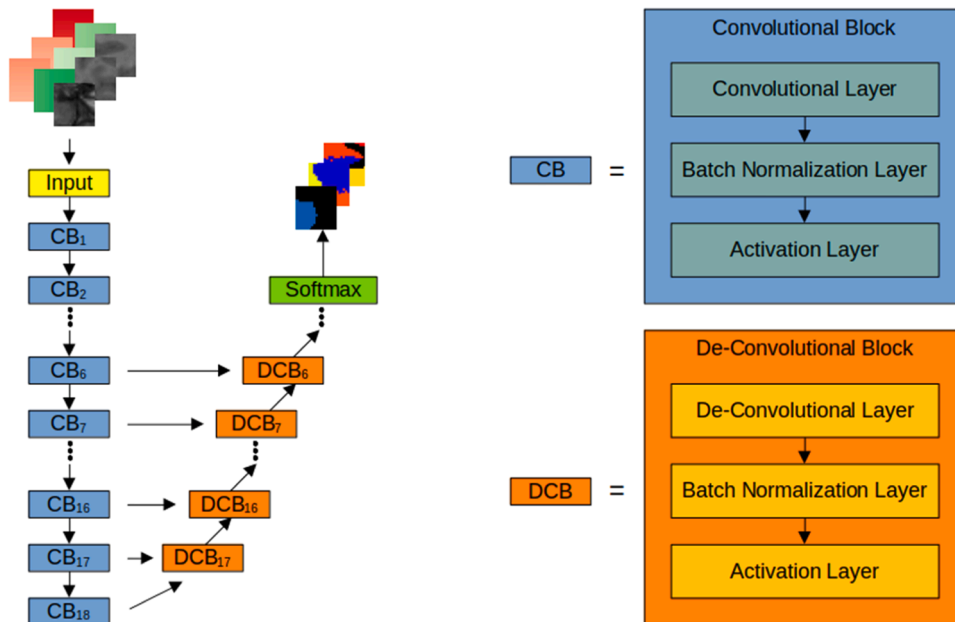
2.3. Training

For training, we break up the 101 Mindboggle scans into randomly selected groups of 81, 10, and 10 for training, validating, and testing respectively. The overlapping window strategy produces 104,385 training windows from the 81 training images and 13,438 validating windows from the 10 validating images. The back propagation algorithm is applied to the training group while the validation group is used only to monitor convergence. Convergence is declared when the validation loss has not decreased in the last 10 epochs.

Fig. 5 shows the time evolution of the training and validating Combo Loss, Soft Dice Loss, and Sparse Categorical Cross Entropy Loss. Minimum Combo Loss occurs at epoch 47, and convergence is declared after training fails to find a new minimum by epoch 57. Fig. 6 shows the time evolution of the weights that contribute to the Combo Loss. The SCCE has a brief initial dominance before rapidly falling off, spending much of the time below 20 %, and from then on serving mainly as a stabilizing term to the gradient calculations during back-propagation. Since the process uses only a single "small-but-deep" neural net to sweep the 3D image, training took only about 30 h on an Nvidia GeForce RTX 3070 Ti, satisfying the requirement for a tool that can be rapidly retrained as needed.

Additional neural nets were trained in the same fashion in order to explore a subset of the parameter space. In addition to the above described net ('AutoParc'), we trained a smaller net that uses an input window of 33 × 33 × 33 and omits the first three convolutional layers ('AutoParc-Small'), a net that uses the Dice Loss only instead of the Combo Loss ('AutoParc-Dice'), and a net that is trained on labels that have not been dilated or had any of the boundaries moved ('AutoParc-Original'). Training and evaluation times were similar for all four.

3. Results and discussion

The remaining ten Mindboggle images that were set aside for testing produce an additional 13,647 sliding windows to evaluate the performance of the model. Evaluating these windows directly through the U-Net to compare to the training and validating values reveals a comparable Combo Loss (training = 0.378, validating = 0.417, testing = 0.398), Soft Dice Metric (training = 0.620, validating = 0.577, testing =

Fig. 4. . AutoParc's U-Net structure, consisting on an Input Block (2D versions of the 3D input images shown for clarity), 18 Convolutional Blocks, 12 De-Convolutional Blocks, and a Softmax output layer. Each Conv/DeConv block consists of a Conv/DeConv layer, a batch normalization layer, and an activation layer.

Table 3

The block dimensions and hyperparameters for the U-Net architecture. The numbering convention of the de-convolutional blocks is such that the sizes match the convolutional blocks.

| Conv. Block #	# of Filters	Output window size	Deconv. Block #	# of filters	Output window size	Hyperparameter	Value
Input	NA	$39 \times 39 \times 39 \times 4$				Kernel Size	$3 \times 3 \times 3$
1	25	$37 \times 37 \times 37$				Activation Fx	ELU
2	25	$35 \times 35 \times 35$				L2 Regularizer	1.00E-06
3	25	$33 \times 33 \times 33$				Batch Size	24
4	25	$31 \times 31 \times 31$				Patience	10
5	25	$29 \times 29 \times 29$	Softmax Output	NA	$27 \times 27 \times 27$		
6	25	$27 \times 27 \times 27$	6	25	$27 \times 27 \times 27$		
7	50	$25 \times 25 \times 25$	7	500	$25 \times 25 \times 25$		
8	50	$23 \times 23 \times 23$	8	500	$23 \times 23 \times 23$		
9	50	$21 \times 21 \times 21$	9	500	$21 \times 21 \times 21$		
10	100	$19 \times 19 \times 19$	10	100	$19 \times 19 \times 19$		
11	100	$17 \times 17 \times 17$	11	100	$17 \times 17 \times 17$		
12	100	$15 \times 15 \times 15$	12	100	$15 \times 15 \times 15$		
13	150	$13 \times 13 \times 13$	13	150	$13 \times 13 \times 13$		
14	150	$11 \times 11 \times 11$	14	150	$11 \times 11 \times 11$		
15	150	$9 \times 9 \times 9$	15	150	$9 \times 9 \times 9$		
16	200	$7 \times 7 \times 7$	16	200	$7 \times 7 \times 7$		
17	200	$5 \times 5 \times 5$	17	200	$5 \times 5 \times 5$		
18	200	$3 \times 3 \times 3$					
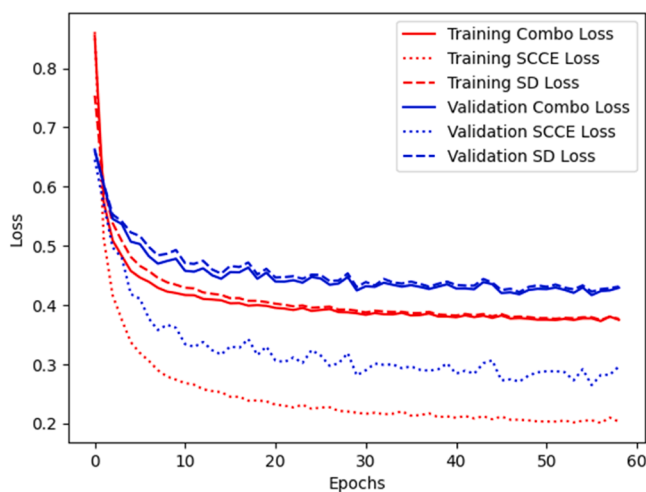
Fig. 5. Time evolution of the training and validating loss and components. The Combo Loss closely resembles the Soft Dice Loss in later epochs, with the SCCE serving mostly as a small stabilizing term. The best model is taken from the point of minimum loss, which occurs at epoch 47.
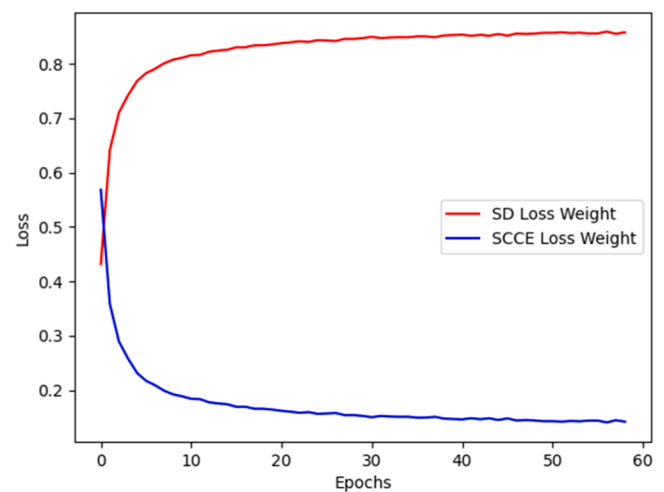

Fig. 6. Time evolution of the individual weights for the SD and SCCE loss that comprise the Combo Loss. The SCCE has a brief initial dominance before serving as a small stabilizing term for the majority of the training.

0.598), and Sparse Categorical Accuracy (training = 0.933, validating = 0.912, testing = 0.924), suggesting that we have sufficient data for evaluation and that the model is sufficiently converged.

For full comparison to the manually labeled images, the ten Mindboggle images are re-assembled from the sliding windows in a roughly reverse fashion to the window generation. Most individual voxels of the full image are represented in several sliding windows. The one-hot encoded value of each of these voxels consists of the average of the one-hot encoded values of its constituent voxels from the sliding windows. A small Gaussian smoother is applied with a half-width of 0.5 voxels in order to reduce any remaining spurious isolated single-voxel predictions (the effect of which is quite small, but serves to remove a few isolated single-voxel predictions in accordance with the view that false positives are more damaging to the eventual use in fMRI overlays than false negatives), and the maximum value of the one-hot encoded is selected as the prediction of the voxel. The Atlas-based FreeSurfer was run on the same 10 testing images for comparison.

Fig. 7 shows a sample comparison of AutoParc vs Manual, AutoParc-Original vs Manual, and FreeSurfer vs Manual for qualitative comparison, showing good overall agreement between both predicted versions
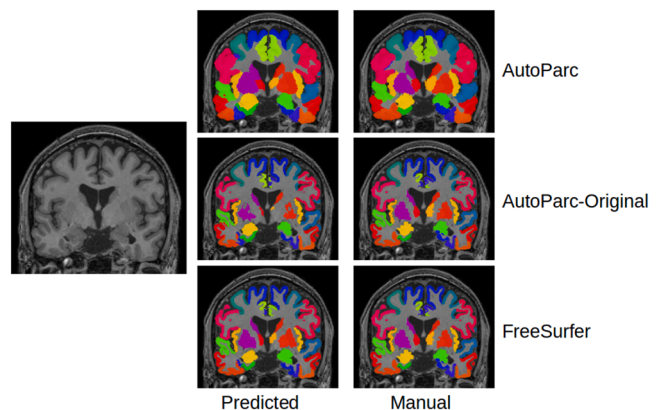

Fig. 7. A qualitative depiction of the performance of AutoParc (top left parcellated image), AutoParc-Original (no dilations or boundary corrections, middle left), and FreeSurfer (bottom left) against the equivalent manually-labeled images (right).

and manual labels, and between equivalent AutoParc-Original and FreeSurfer versions. Fig. 8 shows the quantitative testing results of the four versions of AutoParc. On the x-axis are the labels of the ten testing MindBoggle brains, with the y-axis showing the Dice Coefficient, of the predicted image compared to the manually labeled image. Each box plot represents the Dice Coefficient statistics for all 58 regions for one brain and one AutoParc version.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

The full version of Autoparc has an average Dice Coefficient of 0.810 for the 58 regions and 10 testing images. By comparison, the use of the Combo Loss resulted in only a small improvement in the Dice Coefficient (AutoParc-Dice=0.805) but did result in cleaner training with a smoother training and validating loss history, while the use of a larger T1 window size resulted in a larger improvement (AutoParc-Small=0.789), justifying the increased computational and memory costs associated with the selection.

The largest difference is seen with the version of AutoParc trained on the 58 regions without any of the pre-processing dilations or boundary corrections (AutoParc-Original=0.778). This was an anticipated result, as the internal 'holes' in the defined regions are largely removed via the closing iterations, resulting in more of the predictive information being carried by the more robust normalized orientation data than the fine-scaled T1 intensity. The dilations, both into the white matter and into all 'unknown' voxels, serves both to remove some level of ambiguity in the boundary as well as to decrease the relative size differences between regions, which aids in the calculation of gradients during training. While these changes were made for usage reasons (ie the need to capture the vascular structure in fMRI analysis), it is worth noting that they also significantly improve the performance of AutoParc.

Fig. 9 shows the testing results of each of the 58 regions in AutoParc, with each boxplot showing the statistics for the 10 testing images. The best performing regions, with an average Dice Coefficient of 0.897, are the subcortical structures, which have clearer defined boundaries and thus less ambiguity in the individual labeling samples. The cortical regions have an average Dice Coefficient of 0.800.
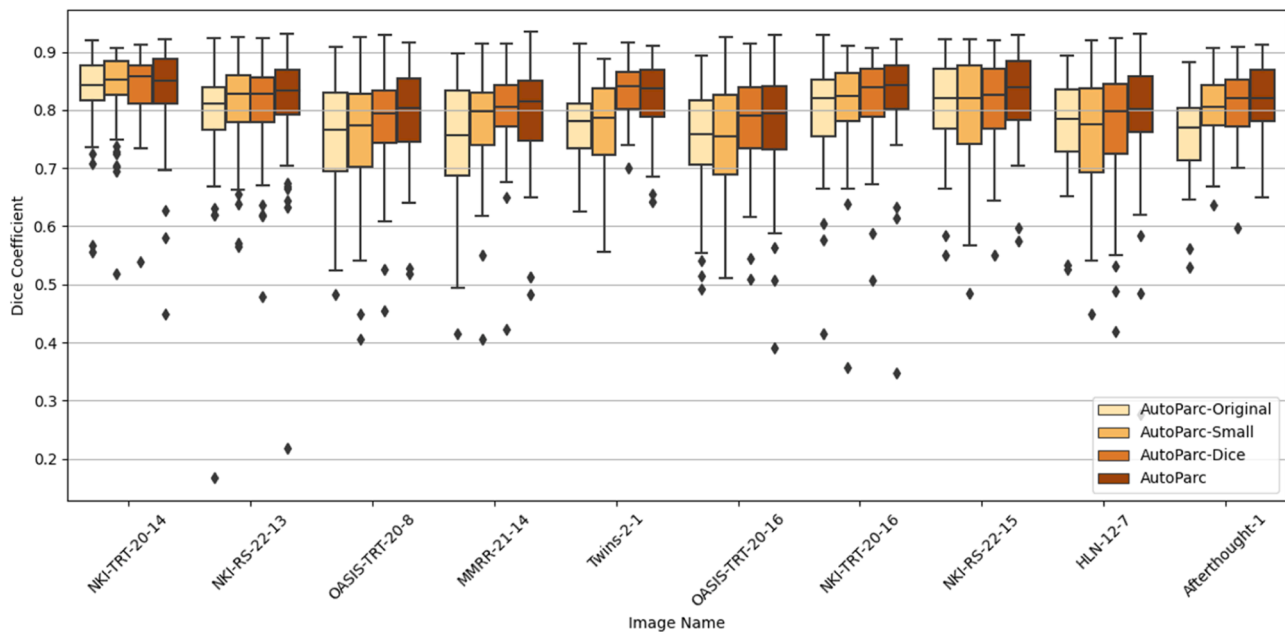
For comparison, the FreeSurfer predictions can be compared to the original version of the manual parcellation that does not include dilations or boundary corrections (Fig. 10). While the average Dice Coefficient is higher than that of AutoParc (0.857, obtained at the expense of a prohibitively long 4.5 h runtime, compared to under 3 min for AutoParc), the overall pattern of high agreement for the less ambiguous subcortical regions and a minority of regions (most notably the left and right pars orbitalis) standing out with low agreement appears similar to that of AutoParc.

A direct comparison of AutoParc and FreeSurfer; Fig. 11, showing the Dice Coefficients comparing AutoParc-Original to FreeSurfer; is telling in that there appears to be a much tighter grouping of the Dice Coefficients around 0.80 despite the fact that AutoParc was trained against the manually-labeled samples and not FreeSurfer generated samples, with a reduction in the number of sub-0.50 dice regions from 7 (AutoParc vs Manual) to 2 (AutoParc-Original vs FreeSurfer). This improved agreement between an existing high-accuracy Atlas-based method and the new modern U-Net method suggests that there is at least a certain amount of human-generated ambiguity in the labeling data that is unlikely to be corrected by further algorithm refinement.

Remaining sources of discrepancy include the tendency of neural networks to smooth small-scale segmentation noise and incorrect protrusions that typically exist in only a few training images (Henschel et al., 2020), resulting in a region prediction that is more accurate than the manual, but has a lower Dice Coefficient. While both this and the other related human-generated ambiguity likely set an upper limit on the achievable Dice accuracy of new codes, the higher performance of FreeSurfer suggests that further algorithm improvement is still possible. Future research will attempt to address this with improved Convolutional-Deconvolutional block architectures, increased training data size (either with new available manual-labeled sample or with FreeSurfer generated samples), increased sliding window size combined with improved processing power, and/or innovative non-cubical sliding window geometries.
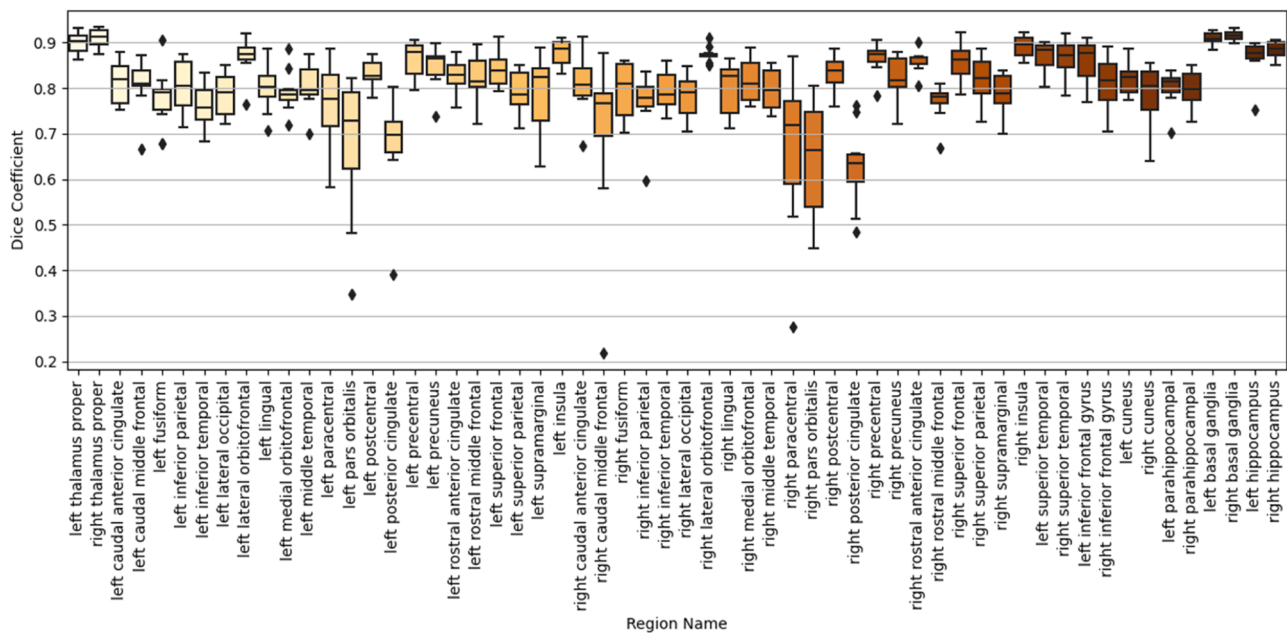

Fig. 8. An illustration of the performance of four version of the AutoParc software on the ten Mindboggle images that were withheld for testing. In addition to the main AutoParc outlined in this paper ("AutoParc", Average Testing Dice Coefficient = 0.810), these include a version using only the Soft Dice loss during training ("AutoParc-Dice", 0.805), a version using a smaller T1 sliding window ("AutoParc-Small", 0.789), and a version trained on the combined labels but without any dilations or inter-region boundary corrections ("AutoParc-Original", 0.778).

Fig. 9. Dice coefficients of AutoParc vs manually-labels for the 10 withheld MindBoggle images. Top performance is seen for the less-ambiguous subcortical regions (0.897). Cortical regions (average 0.800) have general good performance with a small number of poorer performing regions.
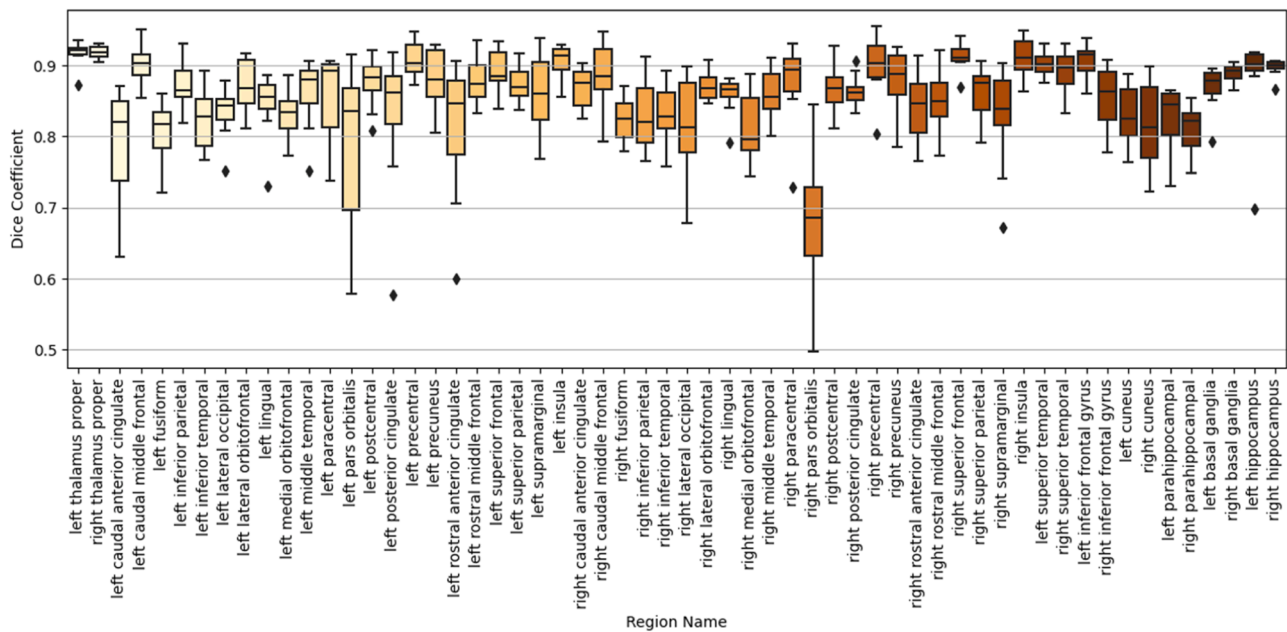

Fig. 10. Dice coefficients of FreeSurfer vs manual for the 10 withheld Mindboggle images.

4. Conclusions

In this paper, we propose a U-Net architecture for parcellation of brain cortical regions and subcortical structures. We demonstrated that a single U-Net applied on a small sliding window is sufficient for the task, provided that it is of sufficient depth and has sufficient filters, allowing for training on tens of thousands of samples obtained from 101 manually labeled samples. The requirement for a tool that can be rapidly retrained based on new learnings was satisfied by a novel loss function that improves the training performance, the need to retrain only the single U-Net, a small enough U-Net to fit on a GPU with a batch size of 24 which increases training stability, and an evaluation procedure that produces the full image via voxel voting. Training the model on dilated

and closed regions (and perhaps re-training as needed) allows the method to be used as the first step in the next generation functional Neuro-Cognitive Imaging tool.

Data statement

The data used in this paper is from the publicly available Mindboggle database at https://mindboggle.info/data

CRediT authorship contribution statement

Heath M. Lorzel: . **Mark D. Allen:** Data curation, Formal analysis, Methodology, Project administration, Resources, Supervision,
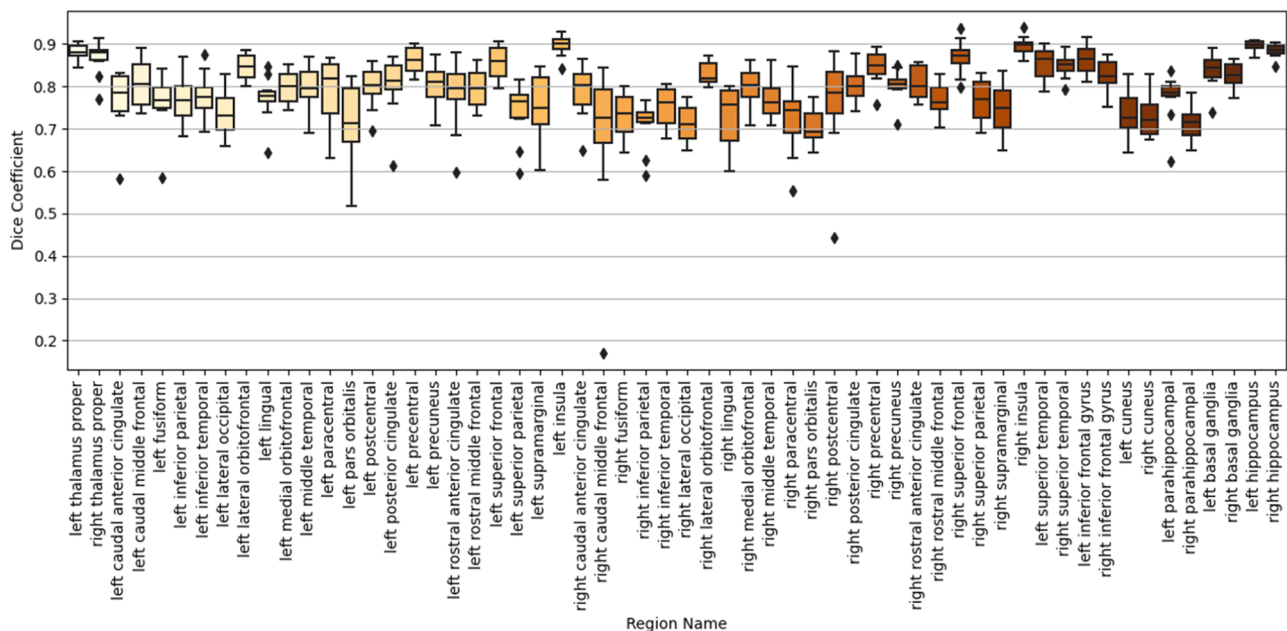
Fig. 11. Dice coefficients of AutoParc-Original vs FreeSurfer for the 10 withheld Mindboggle images.

Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Dr. Heath Lorzel is an employee and Dr. Mark Allen is an owner of Cognitive FX, a neuro-imaging and brain rehabilitation clinic, in Provo, UT.

Data availability

The data is publicly available, while the software is proprietary to Cognitive FX.

References

Allen, M.D., Epps, C.T., 2018. Discovery of therapy-targeting biomarkers for post-concussion syndrome using functional neurocognitive imaging. Brain Disord. Ther. 2018 7, 1.

Beare, R., Lowekamp, B.C., Yaniv, Z., 2018. Image segmentation, registration and characterization in R with SimpleITK. J. Stat. Softw. 86 (8) https://doi.org/10.18637/jss.v086.i08.

Binder, L.M., Rohling, M.L., Larrabee, G.J., 1997. A review of mild head trauma. Part I: meta-analytic review of neuropsychological studies. J. Clin. Exp. Neuropsychol. 19, 421–431.

Cassidy, J.D., Carroll, L.J., Peloso, P.M., Borg, J., von Holst, H., et al., 2004. Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO collaborating center task force on mild traumatic brain injury. J. Rehabil. Med. 36, 28–60.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-net: learning dense volumetric segmentation from sparse annotation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). SpringerCham, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49.

Coupé, P., Mansencal, B., Clément, M., Giraud, R., de Senneville, B.D., Ta, V.T., Lepetit, V., Manjon, J.V., 2020. AssemblyNet: a large ensemble of CNNs for 3D whole brain MRI segmentation. Neuroimage 219, 117026. Oct 1.

Dolz, J., Desrosiers, C., Ayed, I.B., 2018. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. Neuroimage 170, 456–470. Apr 15.

Eierud, C., Craddock, R.C., Fletcher, S., Aulakh, M., 2014. Neuroimaging after mild traumatic brain injury: review and meta-analysis. Neuroimage Clin. 4, 283–294.

Ellis, M.J., Leiter, J., Hall, T., McDonald, P.J., Sawyer, S., et al., 2015. Neuroimaging findings in pediatric sports-related concussion. J. Neurosurg. Pediatr. 16, 241–247.

Epps, C.T., Allen, M.D., 2017. Neurovascular coupling: a unifying theory for post-concussion syndrome treatment and functional neuroimaging. J. Neurol. Neurophysiol. 8 (454), 10–4172.

Faul M.X.L., Wald M.M., Coronado V. (2010) Traumatic brain injury in the United States: emergency department visits, hospitalizations and deaths, 2002-2006. Atlanta, GA.

Fischl, B., 2012. Freesurfer. Neuroimage 62 (2), 774–781.

Giza, C.C., Hovda, D.A., 2001. The neurometabolic cascade of concussion. J. Athl. Train. 36, 228–235.

Giza, C.C., Hovda, D.A., 2014. The new neurometabolic cascade of concussion. Neurosurgery. 75 (Suppl 4), S24–S33.

Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M., 2020. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. Neuroimage 219, 117012. Oct 1.

Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvathaneni, P., Bao, S., Bermudez, C., Resnick, S. M., Cutting, L.E., Landman, B.A., 2019. 3D whole brain segmentation using spatially localized atlas network tiles. Neuroimage 194, 105–119. Jul 1.

Iitzco, DeepBrain, 2018. Available from: https://github.com/iitzco/deepbrain.

Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Hame, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E.C., Keshavan, A., 2017. Mindboggling morphometry of human brains. PLoS Comput. Biol. 13 (3), e1005350 https://doi.org/10.1371/journal.pcbi.1005350.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Leddy, J.J., Kozlowski, K., Fung, M., Pendergast, D.R., Willer, B., 2007. Regulatory and autoregulatory physiological dysfunction as a primary characteristic of post-concussion syndrome: implications for treatment. NeuroRehabilitation. 22, 199–205.

McCrea, M., Guskiewicz, K.M., Marshall, S.W., Barr, W., Randolph, C., et al., 2003. Acute effects and recovery time following concussion in collegiate football players: the NCAA concussion study. JAMa 290, 2556–2563.

McCrory, P., Meeuwisse, W.H., Aubry, M., Cantu, B., Dvorak, J., et al., 2013. In: Consensus statement on concussion in sport: the 4th International Conference on concussion in Sport held in Zurich, 47, pp. 250–258. November 2012. Br J Sports Med.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: In2016 fourth international conference on 3D vision (3DV). Ieee, pp. 565–571. Oct 25.

Mittenberg, W., Canyock, E.M., Condit, D., Patton, C., 2001. Treatment of post-concussion syndrome following mild head injury. J. Clin. Exp. Neuropsychol. 23, 829–836.

Pantazis, D., Joshi, A., Jiang, J., Shattuck, D.W., Bernstein, L.E., Damasio, H., Leahy, R. M., 2010. Comparison of landmark-based and automatic methods for cortical surface registration. Neuroimage 49, 2479–2493.

Rimel, R.W., Giordani, B., Barth, J.T., Boll, T.J., Jane, J.A., 1981. Disability caused by minor head injury. Neurosurgery. 9, 221–228.

Roy, A.G., Conjeti, S., Navab, N., Wachinger, C., 2019. Quicknat: a fully convolutional network for quick and accurate segmentation of neuroanatomy. Neuroimage 186, 713–727.

Thyreau, B., Taki, Y., 2020. Learning a cortical parcellation of the brain robust to the MRI segmentation with convolutional neural networks. Med. Image Anal. 61.

Wachinger, C., Reuter, M., Klein, T., 2018. DeepNAT: deep convolutional neural network for segmenting neuroanatomy. Neuroimage 170, 434–445. Apr 15.

Yu, X., Yang, Q., Zhou, Y., Cai, L.Y., Gao, R., Lee, H.H., Li, T., Bao, S., Xu, Z., Lasko, T.A., Abramson, R.G., 2023. UNesT: local spatial representation learning with hierarchical transformer for efficient medical segmentation. Med. Image Anal 90, 102939. Dec 1.

Zhang, W., Wang, Y., 2019. Geometric brain surface network for brain cortical parcellation. In: InGraph Learning in Medical Imaging: First International Workshop,

GLMI 2019, Held in Conjunction with MICCAI 2019. Shenzhen, China. Springer International Publishing, pp. 120–129. October 17Proceedings 1 2019.

new fMRI protocols. He has also developed prototypes of rehabilitation equipment based on his own experience recovering from a brain injury.

Dr. Heath Lorzel received a PhD in Aerospace Engineering from Arizona State University. His-background includes extensive work in machine-learning, computational analysis, and prototype development in the fields of spacecraft propulsion, hypersonic re-entry vehicles, spin-spray chemical reactors, magnetic materials, metal-air batteries, gas-turbine compressors, hydraulic fracturing, solar power, hydropanels, multi-player games, quantitative investing, and medical imaging. He is currently the Director of Artificial Intelligence at Cognitive FX. His-work includes the development of several machine learning algorithms that automate patient scheduling, brain MRI parcellation, fMRI processing, and analysis of

Mark D. Allen received a PhD in Cognitive Science at Johns Hopkins University with post-doctoral training in Cognitive Neuroscience and Functional Neuroimaging at the University of Washington. Dr. Allen is a co-founder of Cognitive FX, and is a pioneer in the development of fMRI for use in clinical settings. Significant scientific contributions include the development of neuroimaging biomarkers for post-concussion diagnosis and treatment and theoretical work in the pathophysiology of chronic post-concussion symptoms. Specifically, his research led to the discovery of the key role of neurovascular coupling dysfunction in post-concussion syndrome.