U-Net Model for Brain Extraction on Non-human Primates

Xindi Wang[1], Xin-Hui Li[2], Jae Wook Cho[2], Brian Russ[3,4,5], Nanditha Rajamani[2], Alisa Omelchenko[2], Lei Ai[2], Annachiara Korchmaros[2], Pamela Garcia-Saldivar[6], Zheng Wang[7,8], Ned H. Kalin[9], Charles E. Schroeder[3,10], R. Cameron Craddock[11], Andrew S. Fox[12], Alan C. Evans[1], Adam Messinger[13], Michael P. Milham[2,3], Ting Xu[2]

1. Montreal Neurological Institute, McGill University, Montreal, Québec, Canada.
2. The Child Mind Institute, 101 East 56th Street, New York, NY, 10022, USA.
3. Nathan Kline Institute, 140 Old Orangeburg Rd, Orangeburg, NY, USA.
4. Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York City, NY, USA
5.Department of Psychiatry, New York University School of Medicine, New York City, NY, USA
6. Instituto de Neurobiología, Universidad Nacional Autónoma de México Campus Juriquilla, Querétaro, México
7. Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, State Key Laboratory of Neuroscience, CAS Key Laboratory of Primate Neurobiology, Chinese Academy of Science, Shanghai, China
8. University of Chinese Academy of Science, China
9. Department of Psychiatry, University of Wisconsin School of Medicine and Public Health, 6001 Research Park Blvd, Madison, WI 53719
10. Department of Neurological Surgery, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA
11. Department of Diagnostic Medicine, The University of Texas at Austin Dell Medical School
12. Department of Psychology, and the California National Primate Research Center, University of California, Davis, One Shields Ave., Davis, CA 95616
13. Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, USA

*Corresponding Authors: sandywang.rest@gmail.com, ting.xu@childmind.org

Abstract:

Brain extraction (a.k.a. skull stripping) is a fundamental step in the neuroimaging pipeline as it can affect the accuracy of downstream preprocess such as image registration, tissue classification, etc. Most brain extraction tools have been mainly orientated for human data and are often challenging for non-human primates (NHP). In recent attempts to improve the performance in NHP, deep learning models appear to outperform the traditional tools. However, given the minimal sample size of most NHP studies and notable variations in data quality, the deep learning models are very rarely applied in multi-site samples in NHP imaging. To overcome this challenge, we propose to use transfer-learning framework that leverages a large human imaging dataset to pretrain a convolutional neural network (i.e. U-Net Model), and then transferred to NHP data using a small NHP training sample. The resulting transfer-learning model converged faster and achieved more accurate performance than a similar U-Net Model trained exclusively on NHP samples. We improved the generalizability of the model by upgrading the transfer-learned model using additional training datasets from multiple research sites in the Primate Data-Exchange (PRIME-DE) consortium. Our final model outperformed brain extraction routines from popular MRI packages (AFNI, FSL, and FreeSurfer) across multiple heterogeneous multiple sites from PRIME-DE with less computational cost (20s~10min). Our model, code, and the skull-stripped mask repository of 136 macaque monkeys are publicly available for unrestricted use by the neuroimaging community at https://github.com/HumanBrainED/NHP-BrainExtraction.

1. Introduction

As the recent explosion of MRI data sharing in Nonhuman Primate (NHP) scales the amounts and diversity of data available for NHP imaging studies, researchers are having to overcome key challenges in preprocessing, which will otherwise slow the pace of progress (Autio et al., 2020a; Lepage et al., 2020; Messinger et al., 2020; Milham et al., 2018). Among them is one of the fundamental preprocessing steps - brain extraction (also referred to as skull-stripping) (Seidlitz et al., 2018; Tasserie et al., 2020; Zhao et al., 2018). In both human and NHP MRI pipelines, brain extraction is often among the first early preprocessing steps (Esteban et al., 2019; Glasser et al., 2013; Seidlitz et al., 2018; Tasserie et al., 2020; Xu et al., 2015). By removing the non-brain tissue, brain extraction dramatically improves the accuracy of later steps, such as anatomy-based brain registration, pial surface reconstruction, and cross-modality coregistration (e.g., functional MRI, diffusion MRI) (Acosta-Cabronero et al., 2008; Autio et al., 2020a; Lepage et al., 2020; Seidlitz et al., 2018). In humans, automated brain extraction tools have been developed (e.g., the Brain Extraction Tool [BET] in FSL, 3dSkullStrip in AFNI, the Hybrid Watershed Algorithm [HWA] in FreeSurfer, etc.) and easily inserted into a diversity of preprocessing pipelines (e.g. Human Connectome Project [HCP], fMRIPrep, Configurable Pipeline for the Analysis of Connectomes [C-PAC], Connectome Computational System [CCS], Data Processing & Analysis for Brain Imaging [DPABI]) (Cox, 1996; Craddock et al., 2013; Fischl, 2012; Glasser et al., 2013; Jenkinson et al., 2012; Ségonne et al., 2004; Xu et al., 2015; Yan et al., 2016). However, adaption for macaque brain extraction is significantly more challenging, as the data are often noisy due to the smaller brain and voxel sizes involved. The low signal-to-noise ratio (SNR) and strong inhomogeneity of image intensity compromise intensity-based brain extraction approaches, necessitating parameter customization to fit the macaque data (Messinger et al., 2020; Milham et al., 2018). For instance, a new option '-monkey' has been developed to customize AFNI's widely-used 3dSkullStrip function, which improves its performance for NHP data. Yet, the results are still mixed across datasets and often require further manual corrections.

In recent years, registration-based label transferring (i.e. template-driven) approaches have been proposed as a potential solution for NHP brain extraction (Jung et al., 2020; Lohmeier et al., 2019; Seidlitz et al., 2018; Tasserie et al., 2020). These approaches start by registering an individual's anatomical image to the template in order to establish the deformation between the subject-specific head and template head. Once obtained, the transform is used to bring a template-based brain mask back to the individual space, where it can be used to extract the individual brain. The performance of such approaches heavily relies on the accuracy of the transformation and whether the template is representative of the individual data. As such, factors that can compromise the appropriateness or representativeness of the template for the specific individual dataset can decrease the utility of registration-based approaches. Examples where a template may not be representative include variations in species of macaque (i.e. *M. mulatta, M. fascicularis,* etc.) or NHPs (e.g. macaque, marmoset, etc.), the field of view, age (e.g. infant, juvenile, aging), sex (e.g. thicker muscular tissue for male adult macaque), and surgical implants (e.g. with head-holder implants, or anatomical lesions). This issue might be further intensified for data from multiple study centers with different scan acquisitions and samples. In addition, non-linear registration (e.g. ANTs, 3dQwarp) for high-resolution images in brain extraction step is relatively time-consuming (e.g. over hours) and limits the computational efficiency of NHP pipelines.

Recognizing the continued challenges of brain extraction, researchers in human literature have begun to leverage deep learning models as a potential solution. In humans, a growing number of studies have demonstrated the ability of convolutional neural network (CNN) models for brain extraction, as well as tissue segmentation (Henschel et al., 2020; Lyksborg et al., 2015; Rehman et al., 2018; Snehashis Roy et al., 2018; Yogananda et al., 2019). Across studies, training and validation datasets in humans have included hundreds, and in some cases, even thousands of datasets, to ensure accurate performance and avoid overfitting. Once trained, the models have proven to be able to perform highly accurate extraction for new datasets in a matter of seconds. With rare exceptions, the NHP field does not possess datasets close to the multitudes used for training in humans. A recent study did, however, successfully implemented a CNN model (i.e. Bayesian SegNet) for brain extraction using a relatively smaller sample in macaques collected at a single site (N=50), suggesting that such training sets may not need to be as large as expected based upon these preliminary human studies (Zhao et al., 2018). However, the large majority of NHP studies use notably smaller sample sizes (i.e., 2-10). While combining data from multiple studies could be a solution, it is important to note the NHP literature tends to have notably greater variability in imaging protocols than its human counterpart.

The present work attempts to overcome the challenges at hand for NHP imaging by developing a generalizable macaque brain extraction model that can handle data from previously untrained protocols/sites with high accuracy. To accomplish this, we leveraged a transfer learning U-NET strategy, which explicitly aims to train a model for one purpose, and then extend its utility to related problems. In the present case, we trained our model on a human sample (n = 197), and then treated a nonhuman sample (n = 2 for six sites) as the transfer dataset - a strategy that exploits the similarity of human and non-human brain structure. Upon successful demonstration of the ability to transfer between species, we then evaluated the transfer of the updated model to untrained sites in the PRIMatE Data Exchange. Finally, we improved the generalizability of our model by adding a single macaque sample from each of the additional 7 sites (for a total of N=20). We released our pre-trained model, code, and the brain masks outcomes via the PRIMatE Resource Exchange (PRIME-RE) consortium (Messinger et al., 2020; Milham et al., 2018).

2. Methods

2.1. Human Sample

We made use of an open available human brain extraction sample as an initial good-standard training dataset (Puccio et al., 2016). Data were collected as a part of the Enhanced Rockland Sample Neurofeedback Study (N=197, 77 female, age=21-45) (McDonald et al., 2017). Anatomical images data were acquired from a 3T Siemens Trio scanner using a 12 channel head matrix (T1-weighted 3D-MPRAGE sequence, FOV=256x256mm^2, TR=2600ms, TE=3.02ms, TI=900ms, Flip angle=8°, 192 sagittal slices, resolution=1x1x1 mm^3). T1w images were skull-stripped using a semi-automatic iterative procedure that involved skull-stripping all of the data using BEaST (brain extraction based on nonlocal segmentation technique) (Eskildsen et al., 2012) and manually correcting the worst results. Corrected brain masks were added into the BEaST library and the procedure was repeated until the process converged. The results of this procedure underwent an additional manual inspection and correction procedure to identify and fix any remaining errors.

2.2. Macaque Sample

The MRI macaque data used in the present study are publicly available from the recent NHP data-sharing consortium – the non-human PRIMate Data-Exchange (PRIME-DE) (Milham et al. 2018), which includes 136 macaque monkeys from 20 laboratories. We selected one anatomical T1w image per macaque in our analyses. The detailed description of the data acquisition of the magnetization-prepared rapid gradient echo (MPRAGE) image for each site was described in the prior study and PRIME-DE website (https://fcon_1000.projects.nitrc.org/indi/indiPRIME.html, Milham et al., 2018). As the sample size is relatively small in most of the sites (N≤6 for 13 out of 20 sites), we selected six sites which have no less than eight macaque monkeys collected as the first dataset pool for manual edits, model training, and testing (East China Normal University Chen [ecnu-chen], Institute of Neuroscience [ion], Newcastle University Medical School [newcastle], University of Oxford [oxford], Stem Cell and Brain Research Institute (sbri), and the University of California, Davis [ucdavis]). The details of the sample and MRI acquisition are specified in Milham et al ((Milham et al., 2018), Table 1). In total, eight macaque monkeys per site were selected to create the manually-edited 'ground truth' dataset to train and evaluate our model (**Macaque Dataset I**, N=48). To optimize the generalizability of our model to other sites, we also manually edited an additional seven macaques from seven sites (one per site: East China Normal University Kwok [ecnu-k], Lyon Neuroscience Research Center (lyon), Mount Sinai School of Medicine - Siemens scanner [mountsinai-S], National Institute of Mental Health - Messinger [nimh1], Netherlands Institute for Neuroscience [nin], and Rockefeller University [rockefeller]) to create **Macaque Dataset II** (N=7). The rest of the PRIME-DE macaques across all 20 sites were used as an additional hold-out testing dataset (N=81). Of note, our main model was built based on the MPRAGE data. We further made use of the magnetization-prepared two rapid acquisition gradient echoes (MP2RAGE) images from site-UWO-MP2RAGE (N=3) to extend our model to facilitate the brain extraction for MP2RAGE data. All animal procedures were conducted in compliance with the animal care and use policies of the institution where the data was collected.

2.3. Preprocessing

To improve the quality and homogeneity of input data across different sites, a minimal preprocessing was carried out for all anatomical images to remove the salt-and-pepper noise and correct the intensity bias. Specifically, we first re-conformed all T1w images into RPI orientation and applied a spatially adaptive non-local means filtering to remove the 'salt-and-pepper' noise (DenoiseImage in ANTs) (Buades et al., 2011). Next, we performed the bias field correction to normalize image intensities (N4BiasFieldCorrection in ANTs) (Tustison et al., 2010). The preprocessed images were served as inputs for all the brain extraction approaches.

2.4. Traditional Methods and Manually Edited Masks

To compare our deep learning models with state-of-the-art methods for brain extraction, we employed five widely-used skull stripping pipelines implemented in commonly used MRI packages (AFNI, ANTs, FSL, and FreeSurfer) (Avants et al., 2009; Cox, 1996; Fischl, 2012; Jenkinson et al., 2012). Specifically, we tested three intensity-based approaches (FSL BET, FreeSurfer HWA, and AFNI 3dSkullStrip) and two template-driven pipelines (Flirt+ANTS and AFNI @animal_warper) (Jung et al., 2020; Seidlitz et al., 2018; Tustison et al., 2020). The command and parameters of intensity-based approaches were selected based on the experiments and suggestions from the prior studies as follows (Xu et al., 2019; Zhao et al., 2018). 1) FSL 'bet' command with a smaller fractional intensity threshold (-f 0.3) and head radius = 60 mm, 2) FreeSurfer 'mri_watershed' command with default settings, and 3) AFNI '3dSkullStrip' command with

NHP specific option '-monkey' and shrink factor = 0.5. Template-driven approaches in both Flirt+ANTs and AFNI @animal_warper pipelines were performed by first applying a linear registration to transform the individual head to the template head, followed by nonlinear registration. Next, the template brain mask was transformed back into the individual space to obtain the individual brain mask. Specifically, the Flirt+ANTs pipeline uses 'flirt' and symmetric diffeomorphic image registration (SyN) (Avants et al., 2008) for linear and nonlinear registration. Of note, we used FSL 'flirt' rather than ANTS linear registration because 'flirt' is faster and performs better in our initial tests on NHP samples. AFNI @animal_warper uses 3dAllineate and 3dQwarp to compute affine and nonlinear alignments. The same NIMH Macaque Template (NMT) was used in Flirt+ANTs and AFNI @animal_warper pipelines (Jung et al., 2020; Seidlitz et al., 2018). The Macaque Dataset I and II were manually edited by well-trained experts (J.W.C, A.K, and T.X.) using the best output from the traditional approaches as initial masks in ITK-SNAP (http://www.itksnap.org/pmwiki/pmwiki.php)(Yushkevich et al., 2006).

2.5. Train, Update & Evaluation Workflow for Deep Learning Models

2.5.1. Overview

Figure 1 illustrates the overall analytic flow chart of the present study. First, we established a skull-stripping model using the human dataset, aiming to provide an initial pre-trained model to facilitate the transfer-learning from humans to macaques. The human NKI-RS dataset (N=197) was split into a training set (N=190) and a validation set (N=7). We used the training set to train the U-Net model for 10 epochs (i.e. full training dataset has passed through the complete neural network 10 times) and selected the best epoch as the human model based on the performance (i.e. dice coefficient) in the validating set. Next, we transferred the pre-trained human model to build the macaque model using Macaque Dataset I. Specifically, for each of six sites, we randomly selected 2 macaque monkeys as the training set, 1 macaque as the validation set and 5 macaques as the testing set. We also used all macaques across six sites in Macaque Dataset I to create the merged training (N=12), validating (N=6), and testing (N=30) sets. Transfer-learning from human to macaque was carried out for each site as well as the merged data. We calculated 40 transfer-training epochs and selected the epoch that had the best performance in the validation set as the transfer-learning model for each site and the merged samples (refer to the U-Net T model). We also created macaque models that were trained only on the macaque data using the same training (N=12) and validation (N=6) data for each of the six sites and the merged samples (refer to the non-transfer-learning model, i.e. U-Net NoT model). We evaluated and compared the performance between the site-specific T model and the NoT model in testing sets. We also compared the U-Net T and NoT model to the traditional pipelines using the held out test set from Macaque Dataset I. To improve the generalizability of the U-Net model to fit more macaque data from other sites, we further upgraded the U-Net transfer-learning model using both Macaque Dataset I (N=12) & II (N=7) to generate the final generalized model (referred to as generalized U-Net 12+7 model). To evaluate the model performance, we applied the U-Net transfer-learning model, generalized 12+7 model, and traditional pipelines to all the T1-weighted images from PRIME-DE (136 macaques). Expert ratings (details in the Model Evaluation section) were conducted to evaluate whether brain extraction was successful.

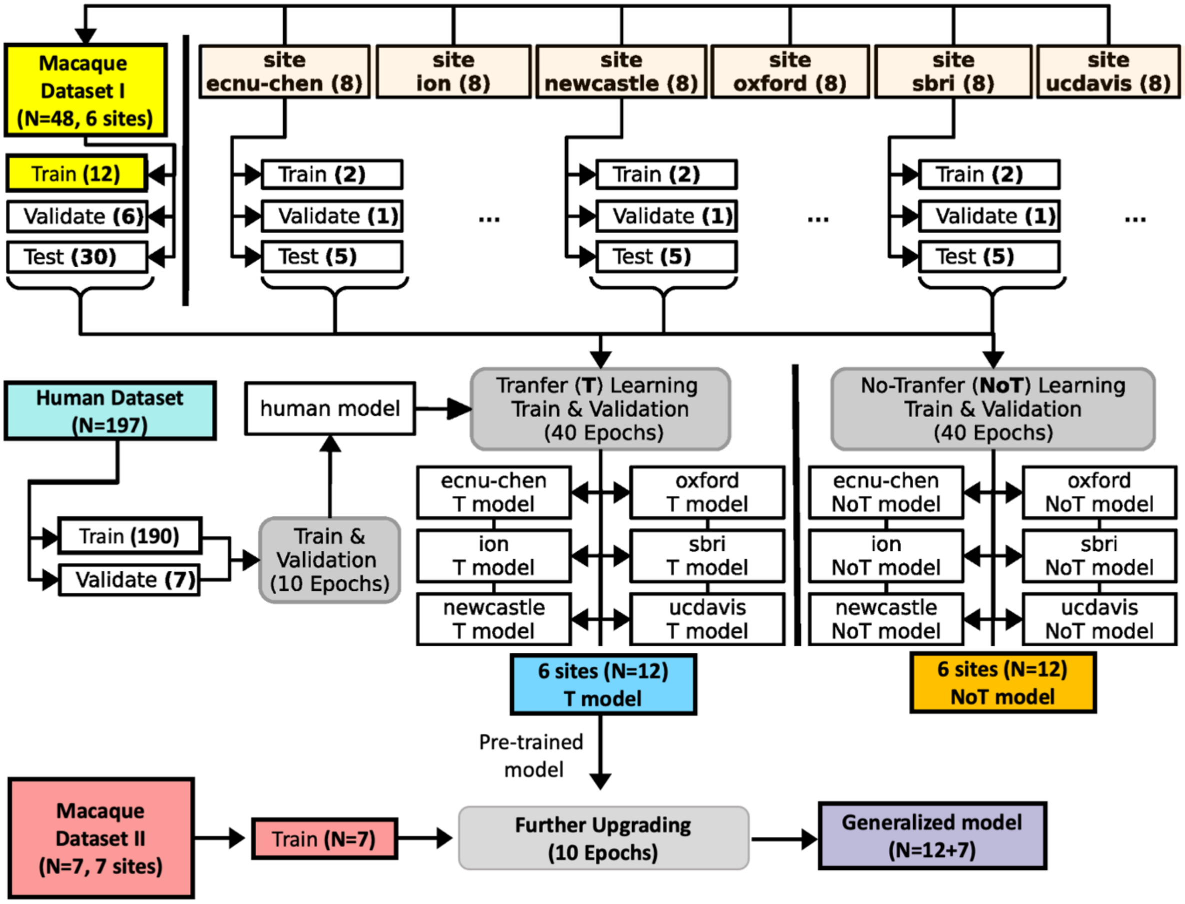

Figure 1. Schematic of the U-Net model training, transfer-learning, and validation on human and macaque datasets. The human dataset was used to generate a pre-train U-Net model for transfer learning. Macaque Dataset I was adopted to train and evaluate the models with and without transfer-learning from the human dataset to small samples of macaque datasets. Further, Macaque Dataset II was added to the training set to generate a generalized model for brain extraction on macaque data.

2.5.2. Neural Network Model (U-Net)

We used a convolutional neural network (CNN) model, U-Net (Ronneberger et al., 2015) for brain extraction. The model was built using an open resource machine learning package (PyTorch: https://pytorch.org). Briefly, the preprocessed 3D T1w images were first resampled into slides along axial, sagittal, and coronal planes. The U-Net model predicts the brain tissue in each slice and then merges all slices to obtain a 3D brain mask. Here, we focused on the architecture of the U-Net model (Fig. S1) and illustrated the details of how the U-Net model identifies the whole brain tissue in the training, validation, and testing processes in the next section.

As shown in Fig. S1, the U-Net model consists of a contraction (i.e. encoding) and an expansion (i.e. decoding) path; each includes five convolution units (Ronneberger et al., 2015). Of note, for a given slice resampled from the T1 images, the input of the U-Net model also included its neighboring slices (i.e. 3 slices in total) as an input (dimension: 3x256x256 blocks). Next, for each convolution unit, two 3×3 convolution layers are built and each is followed by a batch normalization and a leaky rectified linear

(ReLU) operation (Fig. S1, blue arrow). In the encoding step, a 2×2 max pooling with stride 2 was adopted for down-sampling data from the upper unit to the lower unit. We used 16 feature channels in the initial unit, and doubled every unit. The expansive path consists of four up-convolution units. For each up-convolution unit, a 2×2 up-convolution and ReLU operation were applied to the lower unit to yield the feature map for the current unit. This feature map was then concatenated with the feature map at the same level in the contracting path to generate the combined feature map. Similar 3×3 convolution layers with batch normalization and ReLU operation were then performed on the feature maps at each up-convolution unit. Next, we employed a 1×1 convolution layer at the upper un-convolution unit to map the final feature maps to a two-classes map. Finally, a SoftMax layer was used to obtain the probability map for brain tissue. Of note, the initial weights of the convolution and up-convolution layers were randomly selected using a Gaussian distribution $N(0, 0.2)$, and the initial bias of layers was set to 0. For the transfer-learning and model-upgrading model, the parameters from the pre-trained model were used in the initial setting.

2.5.3. Model Training and Validating Procedure

In Figure S2, we illustrated the training procedure of how 3D T1w data was processed for the U-Net model. First, we normalized the intensity of the preprocessed T1w image in the range 0 to 1 across all voxels. We then resampled the T1w volume to a 3D intensity matrix where the highest sampled dimension of T1w volume was forced to be rescaled to 256. In the example shown in Fig. S2, T1w (voxel size: 176x176x96) were rescaled to a 256x256x140 matrix. Next, for each slice along the axial, sagittal, and coronal direction, we generated a 3-slice block; the slice and its two neighboring slices. As a result, we obtained 254, 254, and 138 blocks for axial, coronal, and sagittal directions respectively. Next, we conformed each of the 3-slice blocks into a 3x256x256 matrix. When the dimension of the slice plane was less than 256, we filled the matrix with 0. In total, 646 conformed blocks were generated for the T1w image. Similarly, we processed the manually edited brain mask of the T1w image and generated 646 corresponding blocks. After that, we used the above U-Net model to estimate the probability of brain tissue for each T1w block and calculated the cross-entropy (PyTorch function: CrossEntropyLoss) between the probability map and the 'ground truth' map (Ketkar, 2017) as the model cost. A stochastic optimization (learning rate=0.0001, batch size=20) was then used for backpropagation (Kingma and Ba, 2014).

To evaluate and select the model from the training epochs, we used the probability map generated from the U-Net model to create the final predicted brain mask for each epoch, and then examined the Dice coefficient between the predicted mask and 'ground truth' mask (Fig. S3). Specifically, we processed the validation T1w images into 3-slices blocks following the above procedure. Next, we used the U-Net model at each training epoch to estimate the probability map for each block. All the probability maps (646 blocks) were then combined along the axial, sagittal, and coronal direction and yielded an averaged 3D probability matrix (256x256x256). After that, we rescaled and cropped the matrix back to the original voxel dimension to create a probability volume for the given T1w image. Finally, we thresholded (>0.5) this probability volume to obtain the predicted brain mask. The Dice coefficient between the predicted mask and the 'ground truth' mask was computed for each epoch during training. The epoch which showed the highest Dice coefficient was then selected as our final model.

2.6. Model Evaluation

We carried out a quantitative examination in the testing set of Macaque Dataset I and evaluated the degree to which methods provided more similar brain masks as compared to the manually edited 'ground truth'.

Specifically, we calculated Dice coefficients (Dice) (Sørensen, 1948) between the predicted mask and the manually edited 'ground truth' using the following equation:

$$Dice = \frac{2|P \cap T|}{|P+T|},$$

Where |...| represents the total number of voxels in a mask; *P* and *T* are predicted and 'ground truth' masks, respectively. Dice equals 1 when the predicted mask and 'ground truth' mask are perfectly overlapped. In addition, we also calculated the voxel-wise false-positive (FP), and false-negative (FN) rate to examine where the predicted mask falsely includes the non-brain tissue (higher FP) or misses any brain tissue (higher FN). For each voxel in a given macaque, we tested whether the voxel is falsely assigned as brain tissue (FP=1) or falsely assigned as non-brain tissue (FN=1) in individual space. Next, we transferred the individual FP and FN map to the NMT space by applying affine and warp transforms. Linear affine here was created by aligning the manually skull-stripped brain to the NMT brain ('flirt') and nonlinear warp was generated by registering individual head to the NMT head ('fnirt'). The FP and FN maps were averaged across macaques for each brain extraction approach in the NMT space.

We also employed a qualitative evaluation to compare the success rate across different brain extraction approaches for PRIME-DE data without the 'ground truth'. Three experts with rich experience in imaging quality control visually rated the brain masks (J.W.C., X.L., and T.X.). First, each expert independently reviewed all the images and rated them with four grades, i.e. 'good', 'fair' (the brain tissue was identified with a slightly inaccurate prediction at the edge), 'poor' (i.e. most of the brain regions are identified but with significant errors, in particular missing brain tissue at the edge), and 'bad'. 'Poor' and 'bad' scores were recognized as failed. The brain masks with inconsistent ratings among experts were reviewed and discussed for a final consensus rating.

3. Results

3.1. The Convergence of Human U-Net Model

To acquire a pre-trained model for macaque samples, we first trained a U-Net model using the human samples. Fig. S4 demonstrates the sum of loss on the training set and the mean Dice coefficients across macaques on the validation set for each epoch. After the first epoch, the loss decreased steeply and the mean Dice coefficient reached above 0.985. After that, the mean Dice coefficient gradually improved, showing its highest value (0.9916±0.0012) after the 9th epoch. To avoid over-fitting the human samples for the subsequent macaque training, we only carried out 10 epochs and selected the model after the 9th epoch as the pre-trained model for transfer-learning to the macaque samples.

3.2. Comparison Between Models With and Without Transfer-learning

We first evaluated the site-specific models with and without transfer-learning on Macaque Dataset I. For each site and for the merged sample across the six sites, the loss converged faster for the transfer-learning model than the model without transfer-learning (Fig. 2, left); the loss had nearly reached its minimum after 2 epochs. In addition, the Dice coefficients were more stable for transfer-learning models in the validation sets (Fig. 2, center). Of note, the mean Dice coefficients slowly dropped after 20 epochs in the validation set across three sites (i.e. ecnu-chen, ion, and sbri), which may reflect overfitting of the model selection on small training samples (Fig. 2, center). Nevertheless, the mean Dice coefficients still remained stable in the testing set after the first epoch (Fig. 2, right). In addition, compared to the model trained solely on macaque samples, the transfer-learning model also showed lower variation across testing macaques, suggesting its potential generalizability across macaque samples (Fig. 2, right).

We also computed and evaluated the model with and without transfer-learning based on the merged training samples across six sites (N=12). Similarly, the transfer-learning model (referred to as U-Net T12 model) showed lower loss and higher Dice coefficients across 40 epochs than the model without transfer-learning (Fig. 2, last row); its best performance epoch (i.e. the 37th epoch) was selected as our macaque transfer-learning model (i.e. U-Net T12).

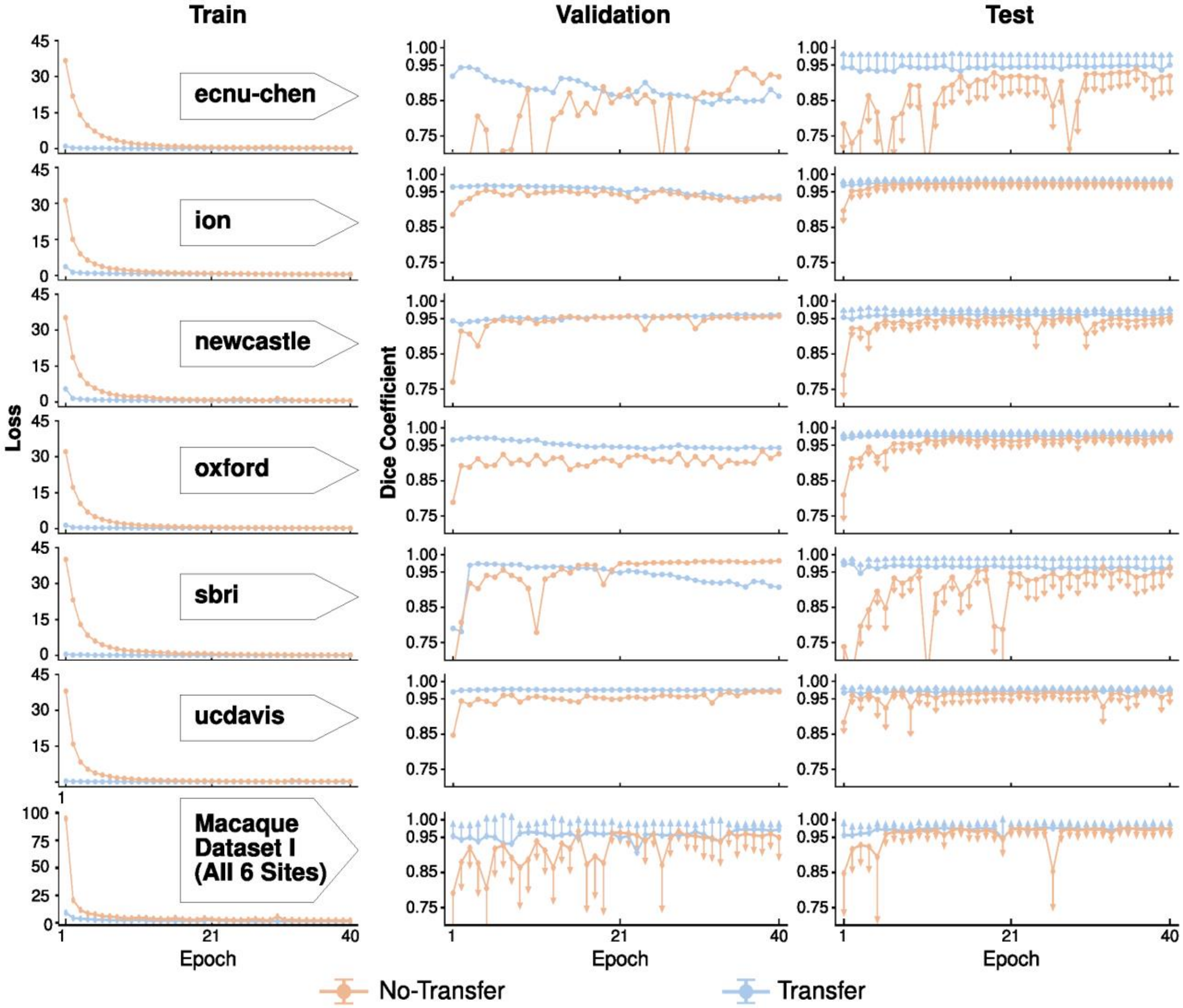

Figure 2. Comparison of skull-stripping performance of the U-Net models with and without transfer-learning on Macaque Dataset I. The models with transfer-learning (blue curve) outperform the one without (orange curve) across each of six sites (row 1-6), as well as the merged sample (the last row). The loss (i.e. the sum of cross-entropy, first column) between the predicted mask and ground truth mask on the train set converges faster for the model with transfer-learning. Similarly, models with transfer-learning show higher Dice coefficients with lower variation in the validation and testing sets than models without transfer-learning. Of note, one-side error bars (standard deviation) were used to avoid dense overlaps between two models.

3.3. Comparison Between the U-Net Model and Traditional Approaches

Recognizing that the model with transfer-learning is superior to the one without, we then updated the transfer-learning model (i.e. U-Net T12) using Macaque Dataset I (N=12) & II (N=7) to generate our final generalized model (U-Net 12+7 model). Here, we evaluated the performance of the U-Net T12 model and the U-Net 12+7 model to other traditional brain extraction approaches. Brain masks from the two U-Net models showed significantly higher Dice coefficients than those from the traditional pipelines (Fig. 3, F=38.531, p<10^{-30} repeated ANOVA, all post-hoc p<0.05). Skull-stripping using the U-Net models was successful (Dice>0.95) for all of the testing macaques (N=30) across six sites. Notably, the U-Net 12+7 model showed relatively higher Dice coefficients than the U-Net T12 model (Fig. 3, *paired-t*=3.62, p=0.001), though the additional training samples used in U-Net 12+7 model were not included in the sites where the testing samples were selected from. This indicated the generalizability of model-upgrading across sites. At the voxel level, both U-Net T12 and '12+7' models exhibited fewer false negatives and false positives than traditional pipelines (Fig. 4). The UNet T12 model showed slightly more false positives than false negatives, which indicated that the model tends to include a few non-brain voxels on the edge of the brain mask rather than miss the brain tissue. Overall, these results demonstrated the feasibility of transfer-learning and model-upgrading using small training samples.

Comparing amongst the traditional approaches, the template-driven approaches (i.e. AFNI @animal_warper and Flirt+ANTS), and AFNI 3dSkullStrip with parameters customized for NHP data (3dSkullStrip -monkey) showed better performance than FSL and FreeSurfer (Fig. 3-4). Both template-driven approaches appear to be more conservative and have missed the brain tissue (low false positives and high false negatives). AFNI 3dSkullStrip missed identifying the brain in the superior regions (higher false negatives on the top) and falsely included non-brain voxels in the bottom (Fig. 4).

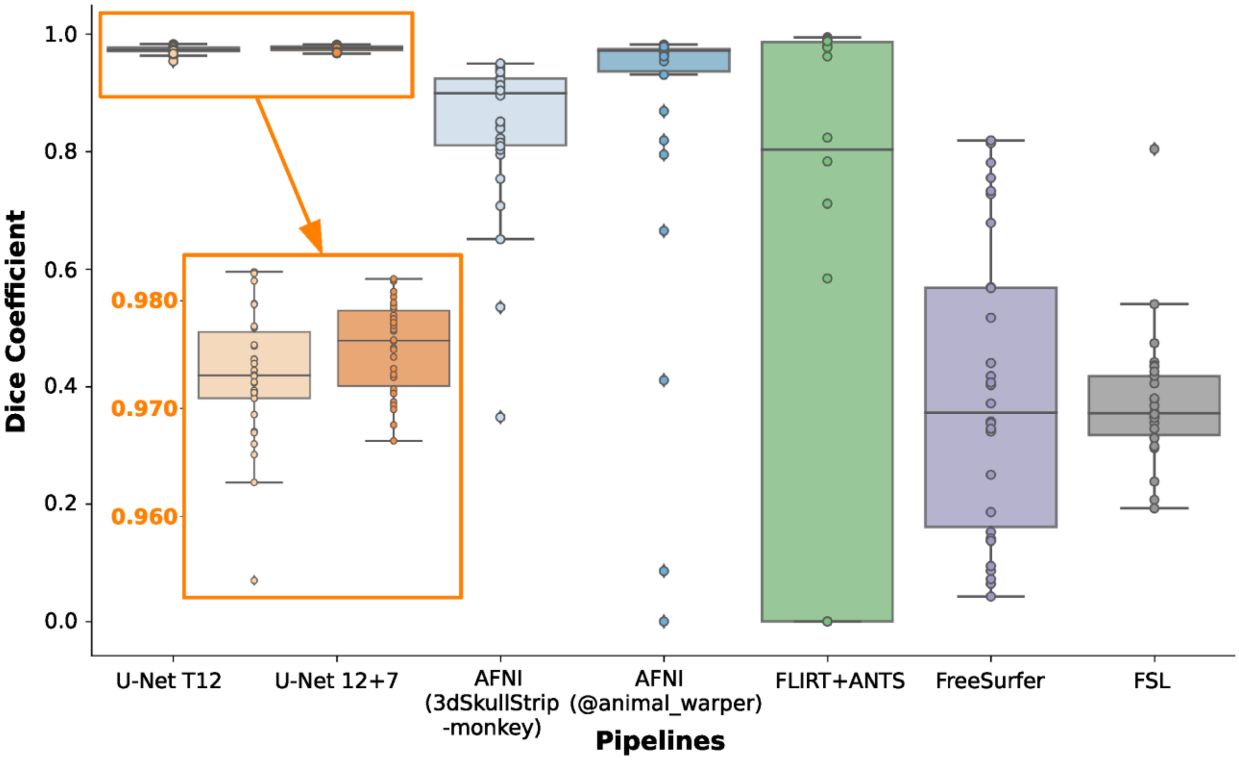

Figure 3. Performance of the U-Net models and traditional approaches. The boxplot shows the Dice coefficients in the testing datasets (N=30) of Macaque Dataset I across brain extraction approaches including 1) the transfer-learning model (i.e. U-Net T12), 2) generalized model (i.e. U-Net 12+7), 3) AFNI 3dSkullStrip command with '-monkey' option, 4) AFNI @animal_warper pipeline, 5) template-driven FLIRT+ANTS pipeline, 6) FreeSurfer HWA approach, and 7) FSL BET approach.

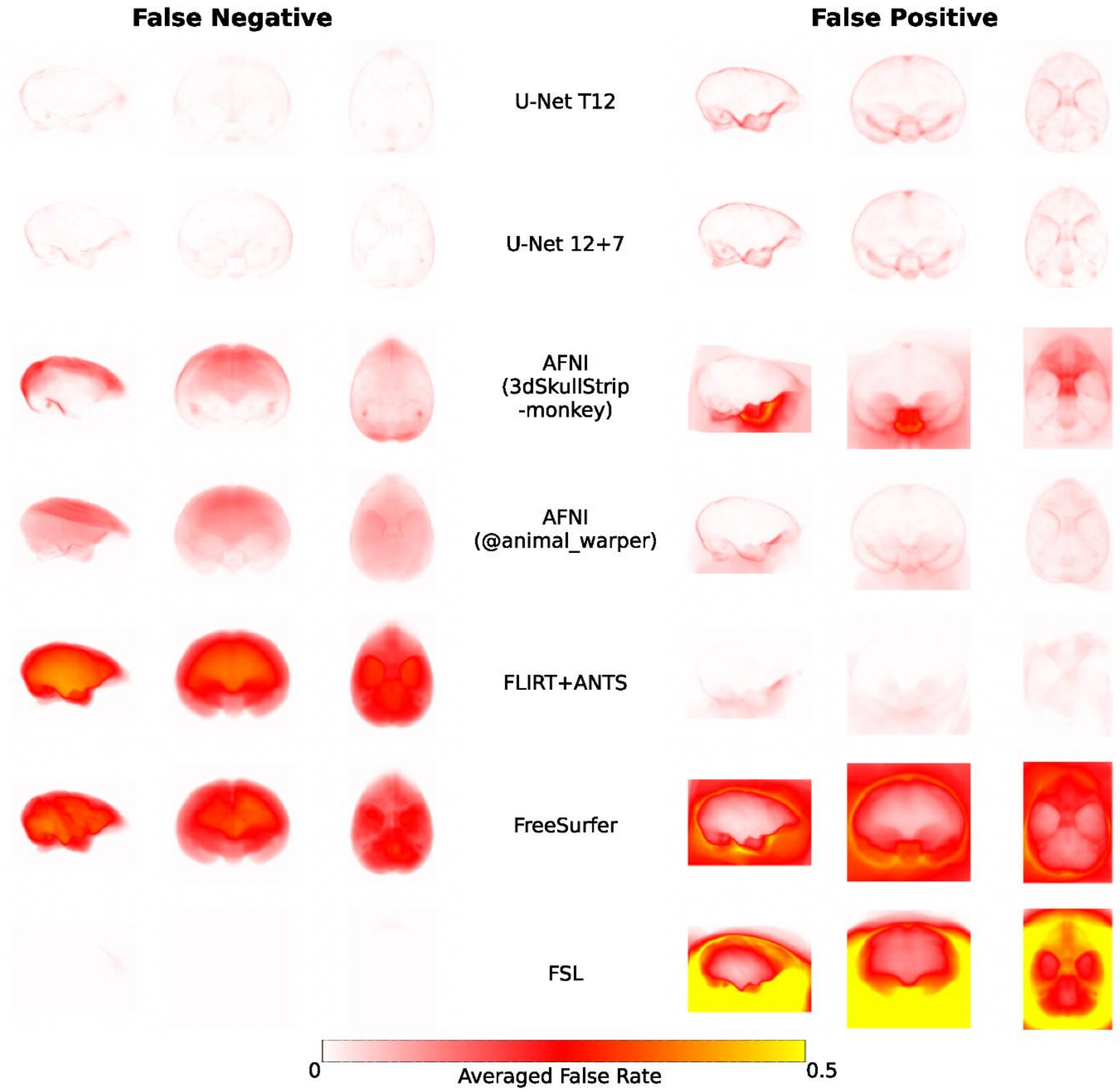

Figure 4. The averaged voxel-wise false-negative and false-positive rates of the U-Net models and traditional approaches in the testing datasets of Macaque Dataset I. The false-negative rate examines where the predicted mask falsely misses the brain tissue in the brain (left column) while the false-positive rate examines where the predicted mask falsely includes the non-brain tissue. Of note, FSL BET tends to falsely include large amounts of non-brain tissue in the brain mask, thus false-negative rates are close to zeros.

3.4. Generalizability of U-Net model and Skull-Stripping PRIME-DE Samples

To evaluate the generalizability of the U-Net model, we applied the U-Net T12 and 12+7 models to all the other macaque samples (i.e. T1w images) contributed to PRIME-DE from differing research sites. The results of brain masks were visually reviewed by three experts and rated into four grades (good, fair, poor, and bad). Good and fair ratings were considered to be successes, while poor and bad ratings were failures. We also used the traditional pipelines and similarly rated their brain mask results. The performance of different pipelines for each macaque is shown on github repository: https://github.com/HumanBrainED/NHP-BrainExtraction. Figure 5 shows the proportion of macaques with good, fair, poor, and bad skull-stripping masks for each of the 20 sites, for each of the five approaches. Again, the U-Net models outperformed the traditional approaches for most sites. In particular, the final U-Net 12+7 model showed the best performance across macaque samples (successful rate=90.4%). All macaque samples (N=123) were successfully skull-stripped across the twelve sites with thirteen exceptions. This result demonstrated the generalizability of the U-Net 12+7 model across sites.

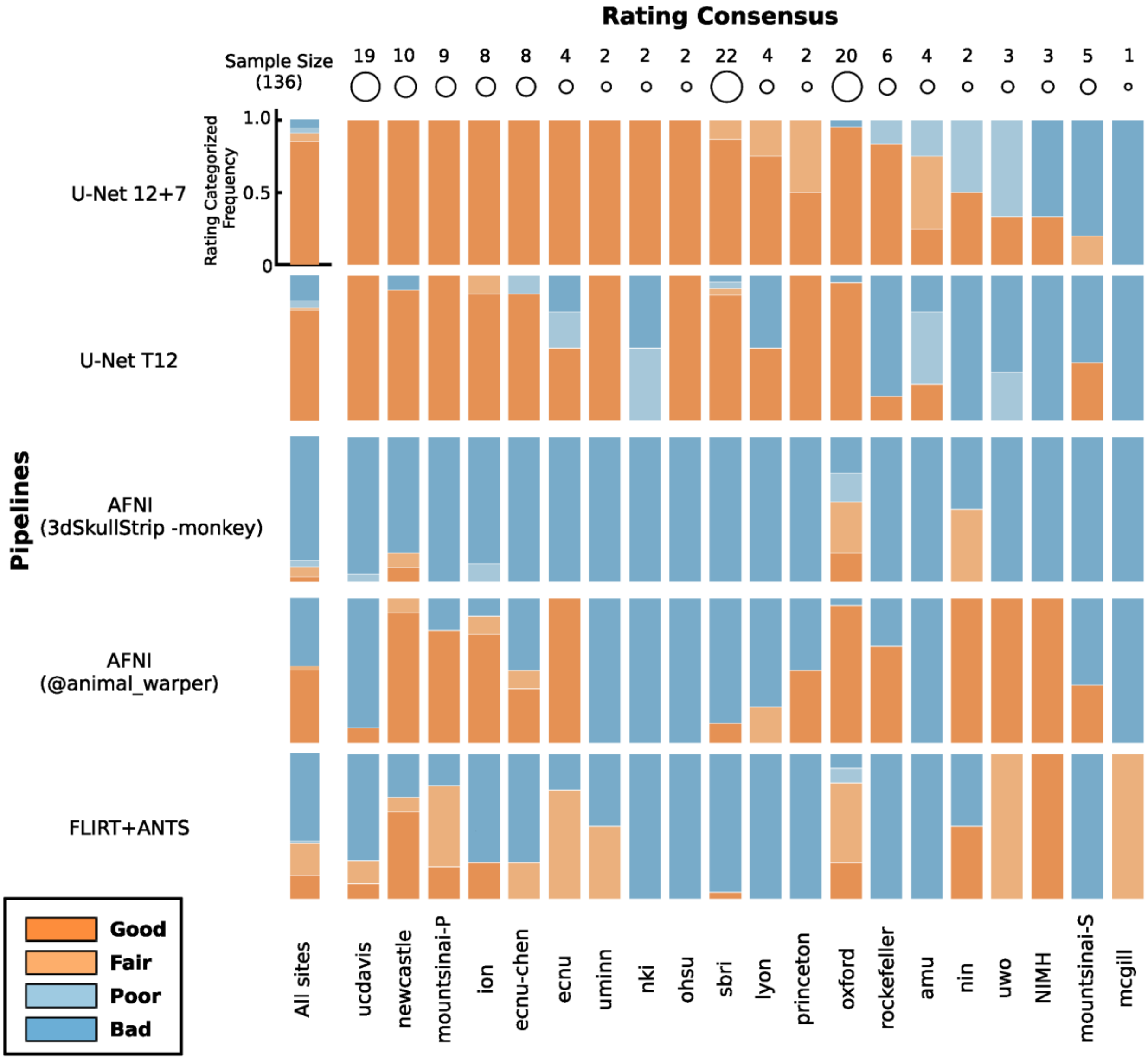

Figure 5. Expert rating consensus of brain extraction performance across the U-Net models and traditional approaches for the PRIME-DE datasets. The stacked bar plots show the proportion of macaques with good, fair, poor, and bad skull-stripping masks on T1w images for each of 20 sites in PRIME-DE. Of note, FreeSurfer HWA and FSL BET default pipelines failed to obtain fair/good brain

masks, and thus their ratings were not displayed. The U-Net 12+7 model (the first row) shows the highest rating across pipelines.

We also noticed that template-driven approaches performed well in some cases, but failed in others (successful rate: AFNI @animal_warper=52.9%, Flirt+ANTS=39.7%). Visual inspection of the intermediate outputs from AFNI @animal_warper and Flirt+ANTS pipelines showed that all successful cases had the first linear alignment of the individual's head to the template head somewhat close. The failures mostly occurred in the first linear registration step; 79.4% of 63 skull-stripping failures from @animal_warper and 89.3% of 84 failures from Flirt+ANTS pipeline had failed in the first linear registration step.

For the failed 13 macaques in the U-Net 12+7 model, when we used the U-Net output mask to perform an initial brain extraction and estimated the affine transformation from the initial skull-stripped brian to the template brain, the template-driven approaches (AFNI @animal_warper and ANTS) turned seven failed cases into successes. We have selected the best successful mask for each macaque and shared all brain masks on PRIME-RE (Messinger et al., 2020).

3.5. Application of U-Net Model-Upgrading for an External Dataset and MP2RAGE Images

Here we used a large external dataset (N=454, UW-Madison dataset with manually drawn whole-brain masks, details described in previous studies) to evaluate the utility of our U-Net -based approach (Fox et al., 2018, 2015; Oler et al., 2010). First, we directly applied the model-prediction module using our U-Net 12+7 model to extract the brain masks for 40 macaques and found relatively good results (Dice=0.923+/-0.025). We further used the U-Net 12+7 model as the pre-trained model and upgraded the U-Net model on these 40 macaques for 10 epochs. This procedure took about 90 min. The upgraded model showed substantial improvement and successfully skull-stripped all remaining macaques (N=414, All Dice>0.95, Mean=0.977±0.005). We further challenged our U-Net model-upgrading module for MP2RAGE images collected from UWO site (N=8) in PRIME-DE, which have different intensity profiles with opposite contrast in gray matter and white matter (Fig. S5). By upgrading the pre-trained U-Net 12+7 model with three MP2RAGE brain masks, the U-Net model enabled skull-stripping on the rest of the hold out MP2RAGE images (N=5).

4. Discussion

The present work demonstrated the feasibility of developing a generalizable brain extraction model for NHP imaging through transfer learning. Central to the success of our effort was our leveraging of human data as a base training set, upon which additional learning for NHP and site-related variations could be readily achieved. We employed the heterogenous, multi-site PRIME-DE resource to evaluate the effectiveness of our framework, finding that the transfer-learning U-Net model identified macaque brain tissue with higher accuracy than traditional methods and also proved more generalizable across data collections. Notably, the transfer-learning U-Net model provides a notably faster solution (approximately 1-10 min) than the next best performing algorithms, which tend to rely on template-based strategies and require over an hour. We have released our model and code, including the utilities for skull-stripping, model-training, transfer-learning, as well as model-updating modules (https://github.com/HumanBrainED/NHP-BrainExtraction). Additionally, we created and shared the skull-stripped repository of 136 macaque monkeys to facilitate the large-scale macaque MRI imaging for the PRIME-DE and NHP community (Messinger et al., 2020; Milham et al., 2018).

A priori, the major roadblocks that one would anticipate for the application of deep learning in NHP imaging are the small sample sizes and variations in imaging protocols (Autio et al., 2020b; Milham et al., 2018). In part, this is a reflection of the experiences of the human imaging community, where large sample sizes have been required to accurately segment the brain using deep learning (Henschel et al., 2020; Kleesiek et al., 2016). The superiority of the transfer learning models in the present work emphasizes the unique advantages of transfer learning in overcoming such challenges for nonhuman primate imaging and offers a model that may be considered in future efforts to overcome similar obstacles in other imaging populations (e.g., macaques with surgical implants and/or in-dwelling electrodes, brain extraction in other species, pathologic models, early development, aging) (Pontes-Filho et al., 2019; S. Roy et al., 2018; Salehi et al., 2018). The success of the transfer-learning model emphasizes the similarity of the general structure of brain tissue in both species (e.g. gray matter, white matter) - despite the anatomical differences between human and macaque heads (e.g. head size, muscular tissue surrounding the skull, skull thickness, etc.) (Yosinski et al., 2014). Aside from accuracy, the transfer-learning model also converged faster (leveling off after 2 epochs) and yielded more stable results across epochs, even though the pre-trained model was established using a different species.

The success of transfer learning in the present work may signal the ability to use smaller samples than previously employed for human imaging studies (Ghafoorian et al., 2017). However, this is not necessarily the case, as this may instead reflect that the folded surface of the macaque is much less complex than that of humans. The macaque central sulcus is less meandering on the lateral parietal lobe (Hopkins et al., 2014). There is only one superior temporal sulcus in the temporal lobe and two less curved sulci (i.e. rectus, arcuate) in the frontal lobe, which results in relatively smooth surface edges for brain extraction (Bogart et al., 2012; Hopkins, 2018). In addition, such folded and meandering brain morphology is substantially more similar across individual macaques than humans (Croxson et al., 2018). As such, in comparison to human brain extraction, the macaque model is feasible using small training samples. Future work in the human imaging community would benefit from a systematic examination of minimal sample sizes needed for successful training and generalization.

Beyond transfer across species, a key finding of the present work is the ability to improve model generalizability across independent imaging sites relative to traditional methods. By further upgrading the pre-trained transfer-learning model based on the secondary training samples across multiple datasets, the upgraded U-Net model has improved the brain extraction performance and showed a higher successful rate than the traditional methods across multiple sites. Of note, the upgraded model enabled successfully skull-stripping datasets acquired from three additional sites that were not included in any training sites. This demonstrates the out-of-site generalizability of upgrading the pre-trained U-Net model across sites. More impressively, we found that the model-upgrading module offers a solution of generalizing the pre-trained model to other modalities (i.e. MP2RAGE), which is usually difficult to achieve with traditional methods. These findings highlight the important role of pre-trained models in brain extraction for small samples, regardless of site (across or within sites), modality (MPRAGE or MP2RAGE), and species (across or within species). Further improvement for the user-specific dataset can be achieved by adding small samples (N≥2) using our U-Net result as the pre-trained model.

It is worth noting that the U-Net model we used in the current study is a 2D convolutional neural network. Although the 3D model usually tends to have higher accuracy (Hwang et al., 2019), the 2D model has a smaller network size, much less memory cost, and requires less computational time. More importantly, the 2D model is more accessible in general computational platforms for users without large amounts of video memory, which can be costly. In addition, unlike the single slice 2D model, our 2D model leveraged the local interslice features by using each slice and its neighboring slices (i.e. 3 slices in total as opposed to just one) as the input image in the first layer. We also resampled the slices in axial, sagittal, and coronal planes and combined the predictions from all three planes into a final probability map (Lyksborg et al., 2015). By doing so, we effectively tripled the number of training slices, which is especially useful for optimizing prediction when training sample sizes are limited (i.e. data augmentation technique). In future work, a conditional random field can be considered in the prediction layer for further refining the weights in the tissue prediction (Chen et al., 2019; Zhao et al., 2018). Beyond improving brain extraction, future efforts may place a greater focus on tissue classification (e.g. GM, WM, subcortical structures). Central to such efforts will be the sharing and amassing of manually segmented brain images, to which 3D CNNs can be applied.

To promote pipeline development in the NHP field, we have released the skull-stripped brain masks, our generalized model, and code via the PRIMate Resource Exchange platform (PRIME-RE: https://prime-re.github.io) (Messinger et al., 2020). Researchers can access the code and perform the brain extraction on their own macaque datasets. The model-prediction for a dataset takes about 20 seconds on a GTX1070 GPU with 700MB GPU memory or 2-10 min on a single CPU core. We also included the model-building and model-upgrading modules in the code, which has been implemented in a recent version of a Configurable Pipeline for the Analysis of Connectomes (C-PAC v1.6: https://github.com/FCP-INDI/C-PAC/releases/tag/v1.6.0). When the results from the current model need improvement, users can upgrade the U-Net using our current generalized model as a pre-trained model and expect a stabilized solution after several training epochs (N<10 epochs) without validation datasets. The model-upgrading module takes about 1-5 hours on a single CPU, or 20 min on a GPU. In addition, we released our manually skull-stripped masks (40 macaques across 7 sites) which can be used as ‘gold standards’ for other deep-learning algorithms. Additionally, we included the successful brain mask outputs to facilitate the further preprocessing analysis of PRIME-DE data.

There are some limitations in our studies. First, although the final U-Net model showed better generalizability across research centers than traditional approaches, it is unable to accurately skull strip all macaque datasets (success rate: 90.4%). This is probably caused by notable variation of macaque samples, where the majority of failed skull strips came from. Of note, even in failed cases, it is possible to use the output of the U-Net model as the initial brain mask to create a prior linear transformation for traditional template-driven approaches - a process that can turn most of the failed cases into success. Second, the U-Net model requires denoising and bias correction using traditional approaches prior to the model prediction. Future work may consider leaving this image noise and the bias field information in the image during training of the network to simplify the processing steps and possibly improve performance.

5. Conclusion

In the present work, we proposed and evaluated a fast and stable U-Net based pipeline for brain extraction that exhibited performance superior to traditional approaches in a heterogenous, multisite NHP dataset. We have released the code for brain mask prediction, model-building, and model-updating, as well as macaque brain masks of PRIME-DE data. We hope this open repository of code and brain masks can promote pipeline development in the NHP imaging field and accelerate the pace of investigations.

Acknowledgement

This work was supported by gifts from Joseph P. Healey, Phyllis Green, and Randolph Cowen to the Child Mind Institute, fundings from National Institutes of Health (NIH R01-MH111439 to CES and MPM, R24MH114806 to RCC and MPM and R01MH120482 to AO). This research was also partially supported by fundings from National Institute of Mental Health (HIMH ZIAMH002918 to AM), R01MH121735 to ASF, the California National Primate Research Center (P51OD011107 to ASF), NIH R01MH101555 to RCC, R01MH081884 to NHK, R01MH046729 to NHK, P50MH084051 to NHK, National Nature Science Foundation of China (81571300, 81527901, 31771174 to ZW), and NIH BRAIN Initiative Grant RF1MH117040 to BER.

Reference:

Acosta-Cabronero, J., Williams, G.B., Pereira, J.M.S., Pengas, G., Nestor, P.J., 2008. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. Neuroimage 39, 1654–1665.

Autio, J.A., Glasser, M.F., Ose, T., Donahue, C.J., Bastiani, M., Ohno, M., Kawabata, Y., Urushibata, Y., Murata, K., Nishigori, K., Yamaguchi, M., Hori, Y., Yoshida, A., Go, Y., Coalson, T.S., Jbabdi, S., Sotiropoulos, S.N., Kennedy, H., Smith, S., Van Essen, D.C., Hayashi, T., 2020a. Towards HCP-Style macaque connectomes: 24-Channel 3T multi-array coil, MRI sequences and preprocessing. Neuroimage 215, 116800.

Autio, J.A., Zhu, Q., Li, X., Glasser, M.F., Schwiedrzik, C.M., Fair, D.A., Zimmermann, J., Yacoub, E., Menon, R.S., Van Essen, D.C., Hayashi, T., Russ, B., Vanduffel, W., 2020b. Minimal Specifications for Non-Human Primate MRI: Challenges in Standardizing and Harmonizing Data Collection. arXiv [q-bio.NC].

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41.

Avants, B.B., Tustison, N., Song, G., 2009. Advanced normalization tools (ANTS). Insight J. 2, 1–35.

Bogart, S.L., Mangin, J.-F., Schapiro, S.J., Reamer, L., Bennett, A.J., Pierre, P.J., Hopkins, W.D., 2012. Cortical sulci asymmetries in chimpanzees and macaques: a new look at an old idea. Neuroimage 61, 533–541.

Buades, A., Coll, B., Morel, J.-M., 2011. Non-local means denoising. Image Processing On Line 1, 208–212.

Chen, W., Liu, B., Peng, S., Sun, J., Qiao, X., 2019. S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. https://doi.org/10.1007/978-3-030-11726-9_32

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S.S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., Others, 2013. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Front. Neuroinform. 42.

Croxson, P.L., Forkel, S.J., Cerliani, L., Thiebaut de Schotten, M., 2018. Structural Variability Across the Primate Brain: A Cross-Species Comparison. Cereb. Cortex 28, 3829–3841.

Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Østergaard, L.R., Collins, D.L., Alzheimer's Disease Neuroimaging Initiative, 2012. BEaST: brain extraction based on nonlocal segmentation technique. Neuroimage 59, 2362–2373.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16, 111–116.

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781.

Fox, A.S., Oler, J.A., Birn, R.M., Shackman, A.J., Alexander, A.L., Kalin, N.H., 2018. Functional Connectivity within the Primate Extended Amygdala Is Heritable and Associated with Early-Life Anxious Temperament. J. Neurosci. 38, 7611–7621.

Fox, A.S., Oler, J.A., Shackman, A.J., Shelton, S.E., Raveendran, M., McKay, D.R., Converse, A.K., Alexander, A., Davidson, R.J., Blangero, J., Rogers, J., Kalin, N.H., 2015. Intergenerational neural mediators of early-life anxious temperament. Proc. Natl. Acad. Sci. U. S. A. 112, 9118–9122.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.-E., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., Wells, W.M., 2017. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. https://doi.org/10.1007/978-3-319-66179-7_59

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., WU-Minn HCP Consortium, 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124.

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.

Hopkins, W.D., 2018. Motor and Communicative Correlates of the Inferior Frontal Gyrus (Broca's Area) in Chimpanzees. Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates 153.

Hopkins, W.D., Meguerditchian, A., Coulon, O., Bogart, S., Mangin, J.-F., Sherwood, C.C., Grabowski, M.W., Bennett, A.J., Pierre, P.J., Fears, S., Woods, R., Hof, P.R., Vauclair, J., 2014. Evolution of the central sulcus morphology in primates. Brain Behav. Evol. 84, 19–30.

Hwang, H., Rehman, H.Z.U., Lee, S., 2019. 3D U-Net for Skull Stripping in Brain MRI. Applied Sciences. https://doi.org/10.3390/app9030569

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790.

Jung, B., Taylor, P.A., Seidlitz, J., Sponheim, C., Perkins, P., 2020. A comprehensive macaque fMRI pipeline and hierarchical atlas. BioRxiv.

Ketkar, N., 2017. Introduction to PyTorch. Deep Learning with Python. https://doi.org/10.1007/978-1-4842-2766-4_12

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv [cs.LG].

Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A., 2016. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. Neuroimage 129, 460–469.

Lepage, C., Wagstyl, K., Jung, B., Seidlitz, J., Sponheim, C., 2020. CIVET-Macaque: an automated pipeline for MRI-based cortical surface generation and cortical thickness in macaques. bioRxiv.

Lohmeier, J., Kaneko, T., Hamm, B., Makowski, M.R., Okano, H., 2019. atlasBREX: Automated template-derived brain extraction in animal MRI. Sci. Rep. 9, 12219.

Lyksborg, M., Puonti, O., Agn, M., Larsen, R., 2015. An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation. Image Analysis. https://doi.org/10.1007/978-3-319-19665-7_17

McDonald, A.R., Muraskin, J., Van Dam, N.T., Froehlich, C., Puccio, B., Pellman, J., Bauer, C.C.C., Akeyson, A., Breland, M.M., Calhoun, V.D., Carter, S., Chang, T.P., Gessner, C., Gianonne, A., Giavasis, S., Glass, J., Homann, S., King, M., Kramer, M., Landis, D., Lieval, A., Lisinski, J., Mackay-Brandt, A., Miller, B., Panek, L., Reed, H., Santiago, C., Schoell, E., Sinnig, R., Sital, M., Taverna, E., Tobe, R., Trautman, K., Varghese, B., Walden, L., Wang, R., Waters, A.B., Wood, D.C., Castellanos, F.X., Leventhal, B., Colcombe, S.J., LaConte, S., Milham, M.P., Craddock, R.C., 2017. The real-time fMRI neurofeedback based stratification of Default Network Regulation Neuroimaging data repository. Neuroimage 146, 157–170.

Messinger, A., Sirmpilatze, N., Heuer, K., Loh, K.K., Mars, R., Sein, J., Xu, T., Glen, D., Jung, B., Seidlitz, J., Taylor, P., Toro, R., Garza-Villareal, E., Sponheim, C., Wang, X., Benn, A., Cagna, B., Dadarwal, R., Evrard, H., Garcia-Saldivar, P., Giavasis, S., Hartig, R., Lepage, C., Liu, C., Majka, P., Merchant, H., Milham, M., Rosa, M., Tasserie, J., Uhrig, L., Margulies, D., Klink, P.C., this issue. A collaborative resource platform for non-human primate neuroimaging. NeuroImage. https://doi.org/10.1101/2020.07.31.230185

Milham, M.P., Ai, L., Koo, B., Xu, T., Amiez, C., Balezeau, F., Baxter, M.G., Blezer, E.L.A., Brochier, T., Chen, A., Croxson, P.L., Damatac, C.G., Dehaene, S., Everling, S., Fair, D.A., Fleysher, L., Freiwald, W., Froudist-Walsh, S., Griffiths, T.D., Guedj, C., Hadj-Bouziane, F., Ben Hamed, S., Harel, N., Hiba, B., Jarraya, B., Jung, B., Kastner, S., Klink, P.C., Kwok, S.C., Laland, K.N., Leopold, D.A., Lindenfors, P., Mars, R.B., Menon, R.S., Messinger, A., Meunier, M., Mok, K., Morrison, J.H., Nacef, J., Nagy, J., Rios, M.O., Petkov, C.I., Pinsk, M., Poirier, C., Procyk, E., Rajimehr, R., Reader, S.M., Roelfsema, P.R., Rudko, D.A., Rushworth, M.F.S., Russ, B.E., Sallet, J., Schmid, M.C., Schwiedrzik, C.M., Seidlitz, J., Sein, J., Shmuel, A., Sullivan, E.L., Ungerleider, L., Thiele, A., Todorov, O.S., Tsao, D., Wang, Z., Wilson, C.R.E., Yacoub, E., Ye, F.Q., Zarco, W., Zhou, Y.-D., Margulies, D.S., Schroeder, C.E., 2018. An Open Resource for Non-human Primate Imaging. Neuron 100, 61–74.e2.

Oler, J.A., Fox, A.S., Shelton, S.E., Rogers, J., Dyer, T.D., Davidson, R.J., Shelledy, W., Oakes, T.R., Blangero, J., Kalin, N.H., 2010. Amygdalar and hippocampal substrates of anxious temperament differ in their heritability. Nature 466, 864–868.

Pontes-Filho, S., Dahl, A.G., Nichele, S., Gustavo Borges Moreno, 2019. A deep learning based tool for automatic brain extraction from functional magnetic resonance images in rodents. arXiv [eess.IV].

Puccio, B., Pooley, J.P., Pellman, J.S., Taverna, E.C., Craddock, R.C., 2016. The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data. Gigascience 5, 45.

Rehman, S., Ajmal, H., Farooq, U., Ain, Q.U., Riaz, F., Hassan, A., 2018. Convolutional neural network based image segmentation: a review. Pattern Recognition and Tracking XXIX. https://doi.org/10.1117/12.2304711

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-24574-4_28

Roy, S., Butman, J.A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2018. Multiple Sclerosis Lesion Segmentation from Brain MRI via Fully Convolutional Neural Networks. arXiv [cs.CV].

Roy, S., Knutsen, A., Korotcov, A., Bosomtwi, A., Dardzinski, B., Butman, J.A., Pham, D.L., 2018. A deep learning framework for brain extraction in humans and animals with traumatic brain injury, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 687–691.

Salehi, S.S.M., Hashemi, S.R., Velasco-Annis, C., Ouaalam, A., Estroff, J.A., Erdogmus, D., Warfield, S.K., Gholipour, A., 2018. Real-time automatic fetal brain extraction in fetal MRI by deep learning, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 720–724.

Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.

Seidlitz, J., Sponheim, C., Glen, D., Ye, F.Q., Saleem, K.S., Leopold, D.A., Ungerleider, L., Messinger, A., 2018. A population MRI brain template and analysis tools for the macaque. Neuroimage 170, 121–131.

Sørensen, T., 1948. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons.

Tasserie, J., Grigis, A., Uhrig, L., Dupont, M., Amadon, A., Jarraya, B., 2020. Pypreclin: An automatic pipeline for macaque functional MRI preprocessing. Neuroimage 207, 116353.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320.

Tustison, N.J., Cook, P.A., Holbrook, A.J., Johnson, H.J., Muschelli, J., Devanyi, G.A., Duda, J.T., Das, S.R., Cullen, N.C., Gillen, D.L., Others, 2020. ANTsX: A dynamic ecosystem for quantitative biological and medical imaging. medRxiv.

Xu, T., Sturgeon, D., Ramirez, J.S.B., Froudist-Walsh, S., Margulies, D.S., Schroeder, C.E., Fair, D.A., Milham, M.P., 2019. Interindividual Variability of Functional Connectivity in Awake and Anesthetized Rhesus Macaque Monkeys. Biol Psychiatry Cogn Neurosci Neuroimaging 4, 543–553.

Xu, T., Yang, Z., Jiang, L., Xing, X.-X., Zuo, X.-N., 2015. A Connectome Computation System for discovery science of brain. Sci Bull. Fac. Agric. Kyushu Univ. 60, 86–95.

Yan, C.-G., Wang, X.-D., Zuo, X.-N., Zang, Y.-F., 2016. DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging. Neuroinformatics 14, 339–351.

Yogananda, C.G.B., Wagner, B.C., Murugesan, G.K., Madhuranthakam, A., Maldjian, J.A., 2019. A Deep Learning Pipeline for Automatic Skull Stripping and Brain Segmentation, in: 2019 IEEE 16th

International Symposium on Biomedical Imaging (ISBI 2019). pp. 727–731.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 3320–3328.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116–1128.

Zhao, G., Liu, F., Oler, J.A., Meyerand, M.E., Kalin, N.H., Birn, R.M., 2018. Bayesian convolutional neural network based MRI brain extraction on nonhuman primates. Neuroimage 175, 32–44.

Supplementary Information:

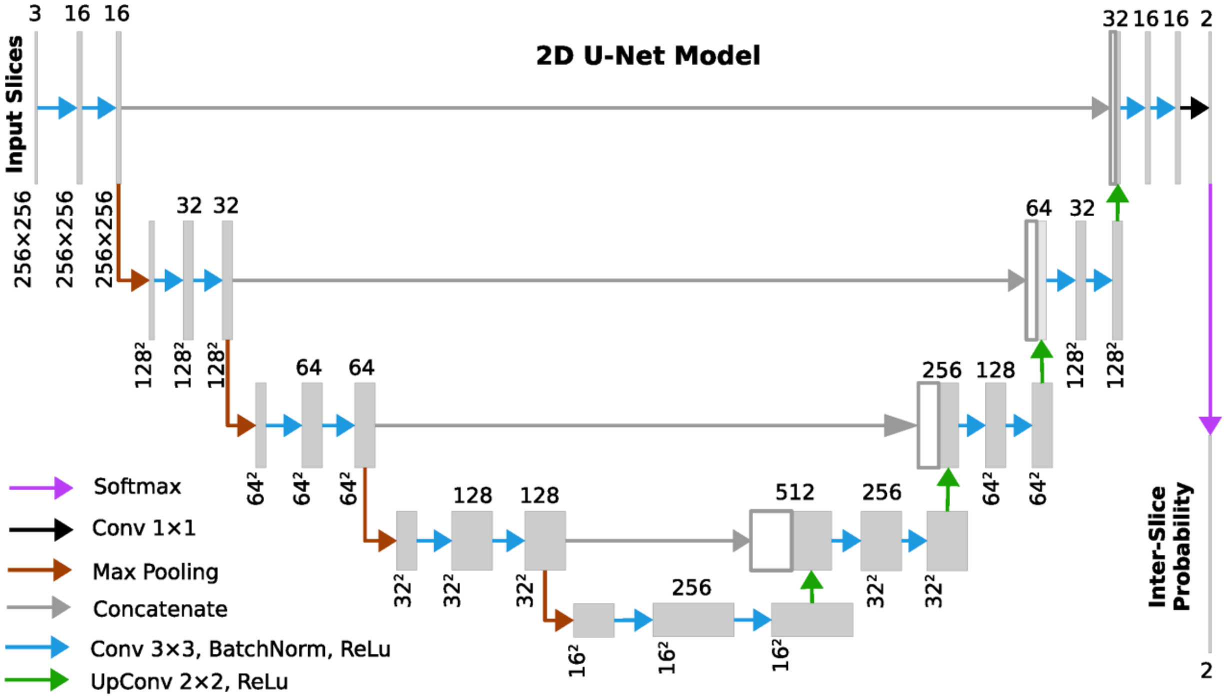

Figure S1. Outline of the architecture of 2D U-Net.

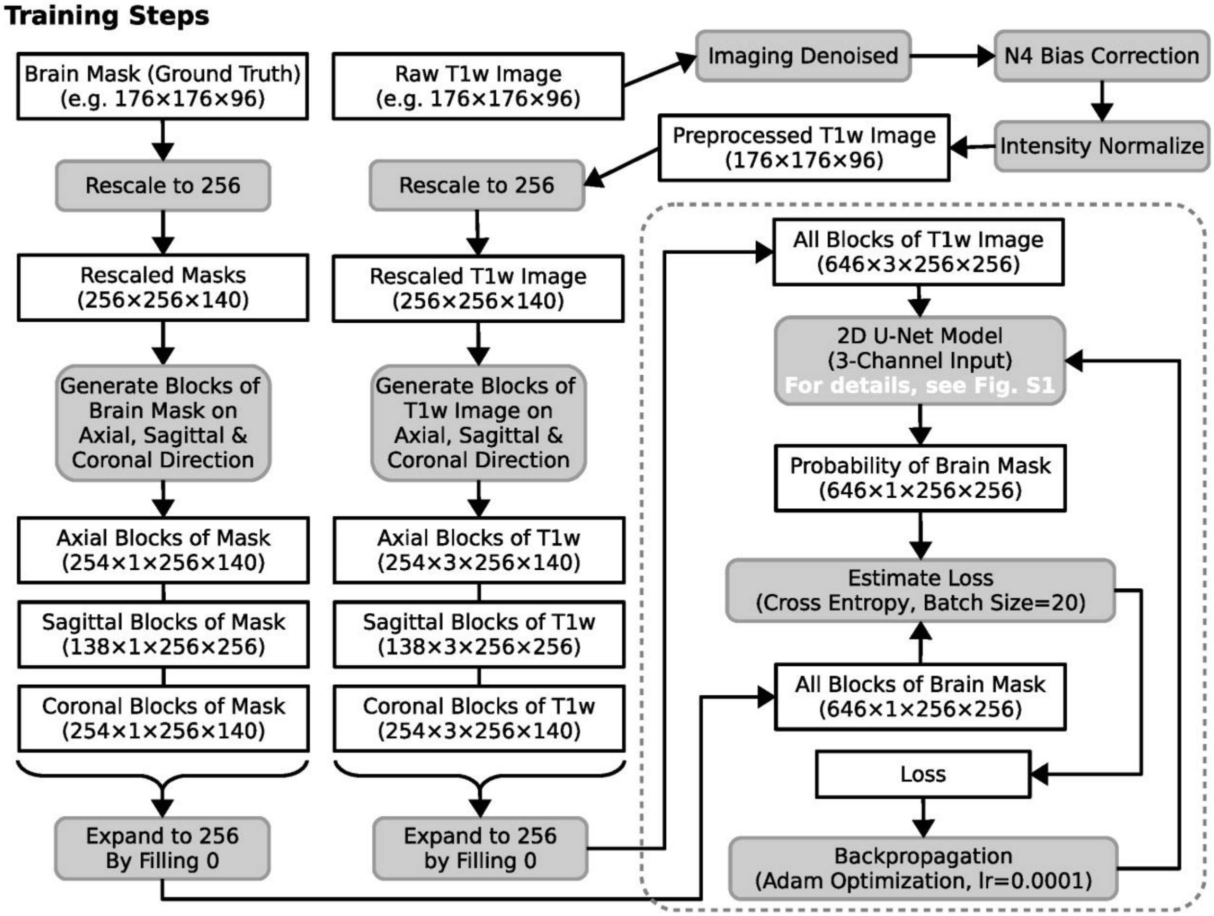
Training Steps
Brain Mask (Ground Truth) (e.g. 176×176×96)
Raw T1w Image (e.g. 176×176×96)
Imaging Denoised
N4 Bias Correction
Intensity Normalize
Preprocessed T1w Image (176×176×96)
Rescale to 256
Rescale to 256
Rescaled Masks (256×256×140)
Rescaled T1w Image (256×256×140)
Generate Blocks of Brain Mask on Axial, Sagittal & Coronal Direction
Generate Blocks of T1w Image on Axial, Sagittal & Coronal Direction
Axial Blocks of Mask (254×1×256×140)
Sagittal Blocks of Mask (138×1×256×256)
Coronal Blocks of Mask (254×1×256×140)
Axial Blocks of T1w (254×3×256×140)
Sagittal Blocks of T1w (138×3×256×256)
Coronal Blocks of T1w (254×3×256×140)
Expand to 256 By Filling 0
Expand to 256 by Filling 0
All Blocks of T1w Image (646×3×256×256)
2D U-Net Model (3-Channel Input)
For details, see Fig. S1
Probability of Brain Mask (646×1×256×256)
Estimate Loss (Cross Entropy, Batch Size=20)
All Blocks of Brain Mask (646×1×256×256)
Loss
Backpropagation (Adam Optimization, lr=0.0001)

Figure S2. U-Net training procedure for each 3D image.

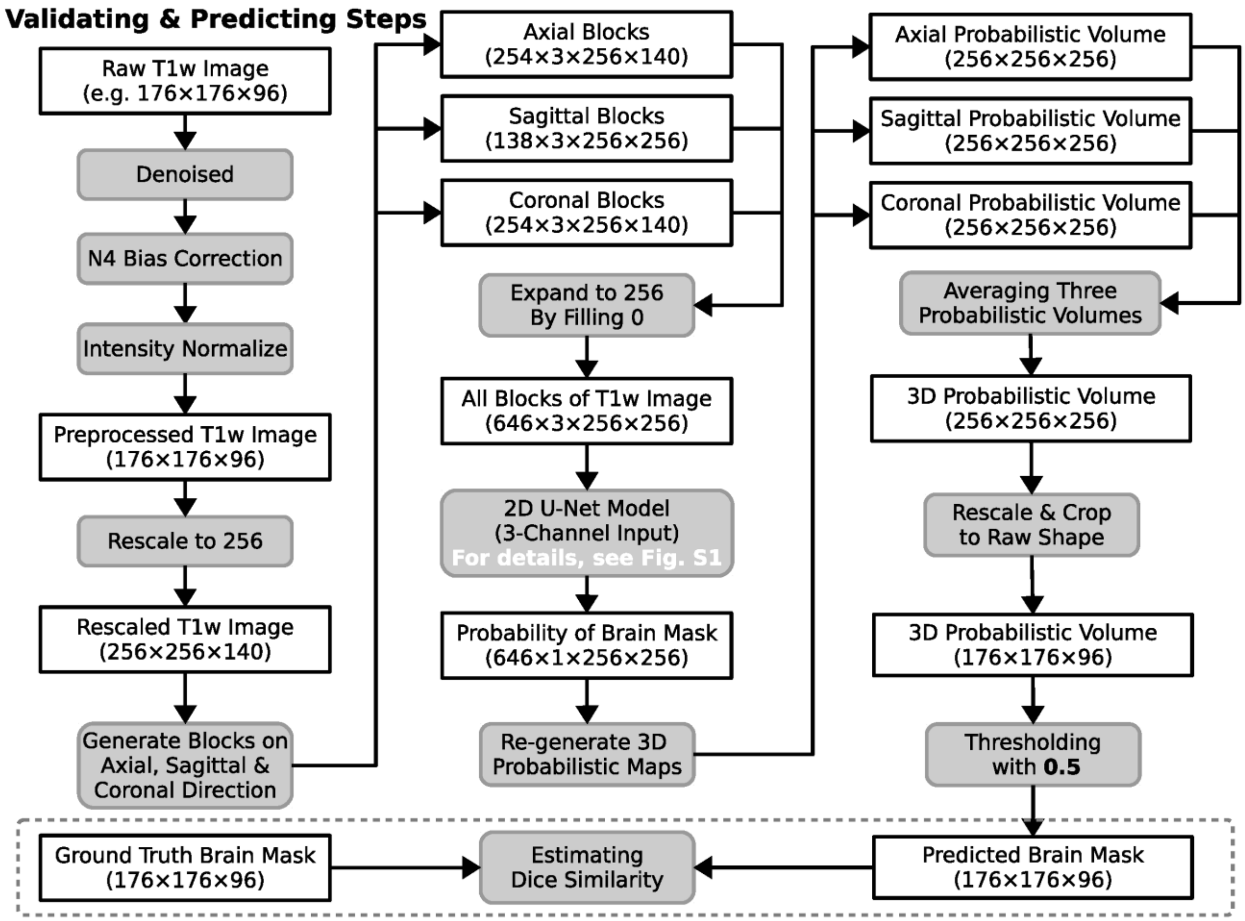
Validating & Predicting Steps
Raw T1w Image (e.g. 176×176×96)
Denoised
N4 Bias Correction
Intensity Normalize
Preprocessed T1w Image (176×176×96)
Rescale to 256
Rescaled T1w Image (256×256×140)
Generate Blocks on Axial, Sagittal & Coronal Direction
Axial Blocks (254×3×256×140)
Sagittal Blocks (138×3×256×256)
Coronal Blocks (254×3×256×140)
Expand to 256 By Filling 0
All Blocks of T1w Image (646×3×256×256)
2D U-Net Model (3-Channel Input)
For details, see Fig. S1
Probability of Brain Mask (646×1×256×256)
Re-generate 3D Probabilistic Maps
Axial Probabilistic Volume (256×256×256)
Sagittal Probabilistic Volume (256×256×256)
Coronal Probabilistic Volume (256×256×256)
Averaging Three Probabilistic Volumes
3D Probabilistic Volume (256×256×256)
Rescale & Crop to Raw Shape
3D Probabilistic Volume (176×176×96)
Thresholding with 0.5
Ground Truth Brain Mask (176×176×96)
Estimating Dice Similarity
Predicted Brain Mask (176×176×96)

Figure S3. U-Net validating and prediction procedure for each 3D image.

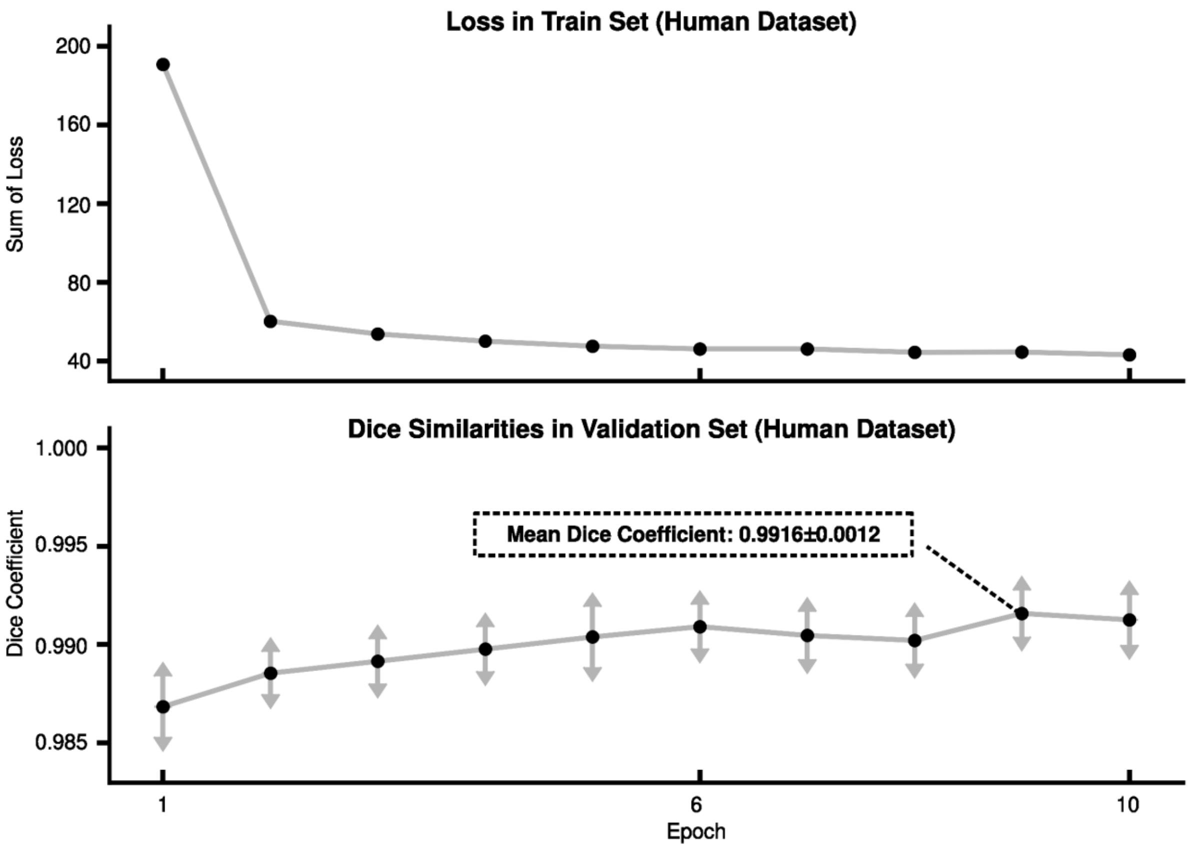

Figure S4. The loss and Dice coefficient of the U-Net model for each training epoch on the human dataset. The mean and standard deviation (arrows) across validation samples was calculated for each epoch.

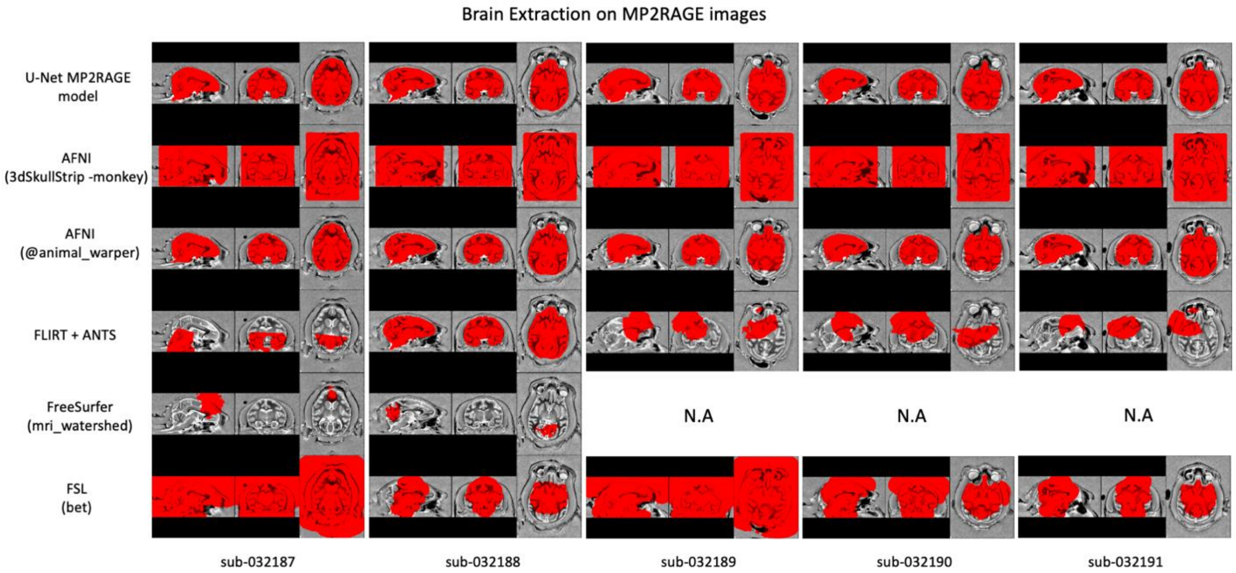

Figure S5. Skullstripping results of U-Net MP2RAGE model and traditional pipelines for MP2RAGE dataset from PRIME-DE (site-uwo).