HHS Public Access

Author manuscript

*Neuroinformatics*. Author manuscript; available in PMC 2020 July 04.

Published in final edited form as:

*Neuroinformatics*. 2018 October ; 16(3-4): 431–444. doi:10.1007/s12021-018-9359-z.

# Field of View Normalization in Multi-Site Brain MRI

**Yangming Ou**[1,2,6,*], **Lilla Zöllei**[2], **Xiao Da**[3], **Kallirroi Retzepi**[2], **Shawn N. Murphy**[4], **Elizabeth R. Gerstner**[5], **Bruce R. Rosen**[2], **P. Ellen Grant**[1], **Jayashree Kalpathy-Cramer**[2,‡], **Randy L. Gollub**[2,6,‡]

[1]Department of Pediatrics and Radiology, Boston Children's Hospital, Harvard Medical School.

[2]Department of Radiology, Massachusetts General Hospital, Harvard Medical School.

[3]Functional Neuroimaging Lab, Brigham and Women's Hospital, Harvard Medical School.

[4]Research Computing, Partners Healthcare.

[5]Neuro-Oncology, Massachusetts General Hospital, Harvard Medical School.

[6]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School.

## Abstract

Multi-site brain MRI analysis is needed in big data neuroimaging studies, but challenging. The challenges lie in almost every analysis step and especially in skull stripping. The diversities in multi-site brain MR images make it difficult to tune parameters specific to subjects or imaging protocols. Alternatively, using constant parameter settings often leads to inaccurate, inconsistent and even failed skull stripping results. One reason is that images scanned at different sites, under different scanners or protocols, and/or by different technicians often have very different fields of view (FOVs). Normalizing FOV is currently done manually or using ad hoc pre-processing steps, which do not always generalize well to multi-site diverse images. In this paper, we show that (a) a generic FOV normalization approach is possible in multi-site diverse images; we show experiments on images acquired from Philips, GE, Siemens scanners, from 1.0T, 1.5T, 3.0T field of strengths, and from subjects 0–90 years of ages; and (b) generic FOV normalization improves skull stripping accuracy and consistency for multiple skull stripping algorithms; we show this effect for 5 skull stripping algorithms including FSL's BET, AFNI's 3dSkullStrip, FreeSurfer's HWA, BrainSuite's BSE, and MASS. We have released our FOV normalization software at http://www.nitrc.org/projects/normalizefov.

### Keywords

Multi-Site MRI; Normalization; Standardization; Field of View

[*]Correspondence: yangming.ou@childrens.harvard.edu.
[‡]Those authors contributed equally.

[6]Information Sharing Statement

The datasets used in this article are mostly from publicly available sources. The software developed in this article is freely distributed at https://www.nitrc.org/projects/normalizefov/

## 1 Introduction

The success and consistency of skull stripping is critical for multi-modal studies investigating structure-function correlation [1,2], for longitudinal studies quantifying neuro-development or brain disease progression [3,4,5,6], and for population studies capturing common phenotypic features of neurological disorders [7,8]. However, multi-site and diverse sets of brain MR images make it difficult to tune skull stripping parameters specific to subjects or imaging protocols. Using fixed skull stripping parameters, however, leads to inaccuracy, inconsistency, and not rarely, failure. One reason is that different MR images may have very different FOVs. Difference in FOVs comes from scanning at different imaging sites, under different imaging protocols or scanners, or by different technicians (e.g., [9,10,11,12,13,14]), or even from subjects across a wide range of ages (e.g., in neurodevelopment studies [11]).

Figure 1 shows how FOV variations can contribute to inaccuracy, inconsistency and even failure in skull stripping. The figure includes MR images of an elderly subject over 60 years of age (Subject A) and of a child 2 years of age (Subject B), scanned in different sites. Images in the original FOV suffer from much larger skull stripping errors (highlighted by red contours in the figure) than images after FOV normalization. This observation is not specific to just one image or just one skull stripping algorithm. Indeed, the effect is common to some widely-used skull stripping methods, including: BET [15,16] and 3dSkullStrip [17], representing deformable surface evolution based methods; BSE [18], representing methods based on edge detection followed by morphological operations; and HWA [19], representing watershed and atlas-based hybrid approaches. Also note that these tools are part of several major neuroimaging analysis pipelines: FSL [20], AFNI [17], BrainSuite [18] or LONI pipeline [21], and FreeSurfer [22], respectively. Therefore, FOV variation is at least one factor contributing to inconsistency, inaccuracy and failure of skull stripping in diverse datasets. This problem has also been recognized in other studies [23,24,25,26].

We hypothesize that it is feasible to automatically normalize FOV without any need for parameter tuning in diverse datasets. We further hypothesize that generic FOV normalization can promote consistency and accuracy of many skull stripping tools in diverse brain MRI datasets. To test these two hypotheses, this paper contains two parts.

**In the first part of this paper**, we propose an atlas-based FOV normalization and validate it in diverse datasets. **In the literature**, FOV normalization is a under-explored topic. Many studies relied on expert manual cropping/truncating of input images [26,27]. The FSL software package provides a ”robustfov” program for image FOV truncation [20]. This program first rotates the 3D image to the orientation of the MNI152 atlas, and then truncates the image, assuming that the brain size in the z dimension (the superior-inferior axes) is 170mm (the default setting). For subjects having different brain sizes, users need to adjust this parameter. The robustfov program was recently used in [28,24]. However, the need for parameter tuning [28] may limit its general use in more diverse datasets. Another recent approach [25] proposed to pre-process images by spatially registering the standard MNI152 atlas into the target image, and then truncating the target image outside the FOV of the warped MNI152 atlas. However, the MNI152 atlas and an arbitrary input image may have

very different FOVs, challenging the registration [29,30]. Similar to [25,20], **our approach** also uses the MNI152 atlas [31] to define a standard FOV. In contrast to them, we use an attribute-based registration algorithm [32] with only its affine component, which has been shown relatively more robust to FOV mismatch [32,30]. Moreover, instead of a single direct MNI152-to-subject registration, we use multiple intermediate templates to build multiple bridged registrations (MNI152-to-template-to-target). We want to emphasize that this is not equivalent to classic multi-atlas segmentation framework, since we only rely on one atlas (MNI152 atlas). This framework is less common, and to our knowledge, the first time it is used for FOV normalization.

**In the second part of this paper**, we tested the effect of FOV normalization on 5 skull stripping algorithms: BET [15,16] from FSL, 3dSkullStrip [17] from AFNI, BSE [18] from BrainSuite and LONI pipeline [21], HWA [19] from FreeSurfer, and MASS [33]. **In the literature**, the effect of FOV normalization on skull stripping has been sometimes mentioned (e.g., [23,24, 25,27]), but mostly with respect to manual FOV truncation, only in one or two skull stripping algorithms/tools (e.g., BET only in [23], multi-atlas skull stripping only in [24]), and moreover, only on one or two relatively homogeneous datasets. Several comparative studies exist evaluating skull stripping algorithms (e.g., [34,35]), but the effect of FOV normalization remains an open question. We also compared the proposed FOV normalization algorithm with one existing FOV cropping tool (FSL's robustfov) for the effects on skull stripping.

## 2 Methods: Part I. FOV Normalization

### 2.1 Overview

Our FOV normalization approach seeks a binary mask. In this binary mask, the foreground identifies the standardized FOV. Two fundamental questions are:

1. How to define the "standard FOV" in an atlas; and
2. How to map the standard FOV from the atlas to an arbitrary image.

The first question is addressed in the first subsection below: Defining a Standard FOV in the MNI152 Atlas. The second issue is addressed in the second subsection: Multiple Bridged Registrations.

Fig. 2 displays the proposed normalizeFOV algorithm. We denoted the MNI152 atlas as $A = (I_A, FOV_A)$, the $N$ intermediate template images $I_{K_n}$, indexed by $n = 1, 2, \ldots, N$, and the target image as $T = (I_T, FOV_T)$. Here, $I$ refers to the intensity image. All images $I_A$, $I_{K_n}$ and $I_T$ are known. $FOV$ is a binary mask whose foreground defines the FOV of the image. $FOV_A$ is known (see Fig. 3). $FOV_T$ is unknown and to be computed. $h$ is the transformation that registers two images. Our goal is to transform the standard $FOV_A$ into the target image space, and to compute $FOV_T$:

$$FOV_T = \text{FUSION}\left(\left\{h_{A \to K_n} \circ h_{K_n \to T}(FOV_A)\right\}_{n=1}^{N}\right) \quad (1)$$

### 2.2 Defining a Standard FOV in MNI152 Atlas

We used the MNI152 atlas [31] because of its inclusion in many state-of-the-art neuroimaging toolkits (e.g., FSL [20], SPM [36] and AFNI [17]) and its use in many morphometry studies (e.g., [37]). As such, we defined the ”standard FOV” as a binary mask, which has value 1 in all the foreground voxels of the MNI152 atlas, with a 5mm dilation margin, and has value 0 everywhere else. This ”standard FOV” binary mask is shown as the blue mask in Fig 3 (overlaid on the MNI152 atlas intensity image). The surface rendering in Fig. 3 allows one to see that, the standard FOV covers the skull, the whole brain, the eyes, and part of the ears and the nose that are 8mm inferior to the most inferior part of the cerebellum, plus a 5mm dilation. The 5mm dilation outside the foreground of the MNI152 atlas was to account for possible transformation errors. We will revisit this choice in the Discussion section.

### 2.3 Multiple Bridged Registrations

An arbitrary input image may have very different FOV from the MNI152 atlas. We recently developed a similarity metric based on texture attributes instead of image intensities. As a result, registration in the presence of FOV mismatch is subject to fewer errors and a largely reduced failure rate [30]. However, it still does not completely eliminate registration failure going from the MNI152 atlas directly to an arbitrary input image. Therefore, we present below a multiple bridged registration framework to replace a single direct MNI152 atlas-to-input registration.

**The Library of Intermediate Templates ($K_n$'s).—**We collected $N = 10$ intermediate templates that had diverse FOVs, scanning sequences, and imaging contrasts. Of these 10 images, 4 were from the Alzheimers Disease Neuroimaging Initiative (ADNI) set [12] (scanned in 3 centers, of which 2 centers used Siemens Trio 3T scanner and the other used GE 3T scanner, 3D magnetization prepared rapid gradient echo (MP-RAGE) sequences, with a typical image size of $200 \times 256 \times 200$ and a typical voxel size of $1 \times 1 \times 1 mm$); 3 were from the Open Access Series of Imaging Studies (OASIS) dataset that was designed for cross-sectional MRI data in young, middle aged, non-demented, and demented older adults (18–96 years old, scanned on a Siemens 1.5T scanner at the University of Washington, MP-RAGE sequence, with a typical image size of 256×256×128 and a typical voxel size of 1×1×1mm) [38]; and 3 were scanned at the University of California, Los Angeles, with a high-resolution 3D Spoiled Gradient Echo (SPGR) sequence on a GE 1.5T scanner, with a typical image size of $256 \times 300 \times 256$ and a typical voxel size of $1 \times 1 \times 1 mm$) [39,40].

Note that, these templates have very different original FOVs (the yellow contours in Fig. 4). They were chosen intentionally, so that the intermediate templates of varying FOVs could potentially better cover a wide variety of FOVs in an arbitrary image. Note that, the fact that templates have different FOVs from that of the MNI152 atlas was not a big concern, since we could ensure that the pre-computed registrations between the MNI152 atlas to the intermediate templates are accurate. Such registrations need to be done once only, and then kept fixed. The registrations between intermediate templates (the bridges) to an arbitrary input image need to be re-computed when a new image is presented.

**Precomputed Registration between the MNI152 Atlas and Intermediate Templates** $\left(h_{A \to K_n}\right)$**.—**We calculated the affine transformation between the MNI152 atlas and each of the templates by registering their skull-stripped versions (skull stripping masks were provided by the public databases). Note that we used skull-stripped brains here to greatly reduce the difficulty of registration. Fig. 4 shows $h_{A \to K_n}{}^{\circ} h_{K_n \to T}(FOV_A)(n = 1, \ldots, N)$ when $K_n = T\left(\text{i.e.}, h_{K_n \to T} = \textbf{Identity}\right)$ as outlined by the red contours. These registrations are pre-computed and kept the same for every input target image.

**Registration between Intermediate Templates and the Target** $\left(h_{K_n \to T}\right)$**.—**We computed $h_{K_n \to T}$ by an attribute-based registration tool that showed increased robustness dealing with FOV mismatch [30] (the "-a 2" argument for robust affine registration in the DRAMMS software). Unlike the pre-computed registrations between the MNI152 atlas and the intermediate templates, where we used skull-stripped images to improve registration accuracy, the registration between intermediate templates to an input target was only computed using the with-skull images. This was because skull-stripping was our final goal and thus not assumed in an arbitrary input image.

**Combining Multiple Bridged Registrations (Fusion(·).)—**Due to FOV mismatch, we could not assume that all bridged registrations were successful. We did assume, though, that at least some succeeded. This was because the intermediate templates covered a wide range of FOVs. A situation in which all intermediate templates failed to align with the target image should be rare, and indeed did not occur in our experiments with very diverse data. The problem, then, was to effectively select the successful bridged registrations. We ranked the quality of the bridged registrations by the similarity with the target image after registration, within the intersection of their FOVs, and selected the intermediate templates by the following criterion:

$$\begin{aligned} \mathrm{SIM}_n &= \mathrm{CC}_{\Omega_\cap}\left(I_T, h_{A \to K_n \to T}(I_A)\right) \\ &\times \mathrm{NMI}_{\Omega_\cap}\left(I_T, h_{A \to K_n \to T}(I_A)\right) \\ &- \left\|\det\left(\mathbf{Id} - h^S_{A \to K_n \to T}\right)\right\| \end{aligned} \tag{2}$$

$$where \quad \Omega_\cap = \Omega_T \cap \Omega_{h_{K_n \to T}(K_n)} \tag{3}$$

$$and \quad h_{A \to K_n \to T} = h_{A \to K_n}{}^{\circ} h_{K_n \to T} \tag{4}$$

The selected templates formed a set $Q_T$,

$$Q_T = \{K_n \mid \mathrm{SIM}_n \geq \alpha M\} \quad \text{where} \quad M = \max_n \mathrm{SIM}_n \tag{5}$$

Here $\alpha$ was a threshold parameter that we empirically set to 0.9. The final normalized FOV was the STAPLE-based [41] fusion of the standard FOV as propagated by those selected templates, i.e.,

$$FOV_T = \text{STAPLE}\left(\left\{h_{A \to K_n \to T}(FOV_A)\right\}_{n=1}^{|Q_T|}\right) \quad for K_n \in Q_T \tag{6}$$

### 2.4 Validations

**Datasets.—**We used T1-weighted MR images from 126 subjects in 5 datasets. As Table 1 shows, the ages of those subjects ranged from birth to 90 years old. Their images presented different FOVs and different image contrasts. They were from at least 5 different imaging sites, with scanners of different vendors and strengths of magnetic fields.

**Qualitative Evaluation of FOV Normalization.—**We applied the whole framework to all 126 images in all 5 testing datasets. We qualitatively examined (a) bridged versus direct registration; and (b) multiple bridged registrations, with the proposed selection criterion (Eqns. 7 and 8) in place, versus single bridged registration. Finally, we visualized the FOV normalization results for some randomly-chosen subjects in each of the 5 datasets.

**Quantitative Evaluation of FOV Normalization.—**We did this in a leave-one-out cross validation framework in the 10 intermediate templates. Each time we picked an intermediate template as the target image, we used the MNI152 atlas and the remaining 9 intermediate templates to find the normalized FOV in the target image. The ground-truth normalized FOVs in the target images were not available. Alternatively, we compared the normalized FOV mask with the transformed MNI152 FOV mask based on directly registering the skull-stripped MNI152 atlas and the skull-stripped target image. In this comparison, we used the Dice Coefficient between the masks for quantitative evaluation of the accuracy of FOV normalization.

## 3 Method: Part II. Effect of FOV Normalization on Skull Stripping

### Skull Stripping Diverse Datasets with and without FOV Normalization.

We measured the accuracy of skull stripping before and after FOV normalization in the 5 diverse datasets. The accuracy was measured by the Dice Overlap Coefficient against manually annotated brain masks. We used 5 skull stripping tools: BET, 3dSkullStrip, BSE, HWA and MASS. They represent a variety of algorithms as we described in the Introduction. As such, we performed a total of 1,260 skull stripping jobs (=126 images $\times$ 2 variants (before and after FOV normalization) $\times$ 5 skull stripping algorithms).

The parameter settings for a skull stripping algorithm was kept fixed before and after FOV normalization. We set the parameters as follows.

For BET, we set ”fractional intensity threshold” (the ‘-f’ option) at 0.3 and the ”vertical gradient of fractional intensity threshold” (the ‘-g’ option) at 0. This setting was the same as in [44] and [27], which independently found the above parameters optimal for skull stripping

a set of adult brain images, and a set of pediatric brain images, respectively. Two other studies set the parameters for BET otherwise — [23] at f=0.1 and g=0, which was close to the parameter setting we adopted, and [45] at f=0.65 and g=−0.3.

For 3dSkullStrip, we followed [45], which examined various parameter combinations and set ”shrink factor” (the ‘-shrink_fac’ option) at 0.7 and set the ”speed of expansion” (the ‘exp_fac’ option) at 0.1.

For BSE, there are two important parameters: the ”edge detection constant” (the ‘-s’ argument) and the ”diffusion constant” (the ‘-d’ argument). Some set the two parameters at 0.7 and 10 [45], or 0.6 and 5 [27], or 0.7 and 20 [44], but the results were bad in our experiments (visual examinations). So we followed [46] to set these two parameters at 0.62 and 35, and another parameter #iteration=3. The same parameters were adopted in other studies [47,48].

For HWA, we adopted the same setting suggested in both [45] and [44], using the ‘-less’ argument while keeping other options default.

For MASS, we used the same setting suggested in [33], which used the default parameter for multi-site images.

### Comparison with another FOV Truncation Tool.

We compared two FOV normalization approaches (FSL’s robustfov and our proposed normalizeFOV), based on how they impacted the consistency, accuracy and success rate of skull stripping, by the aforementioned 4 skull stripping algorithms in 5 datasets.

*Accuracy* was measured by the Dice Coefficient between the manual and algorithm-computed skull stripping results before and after each FOV normalization tool. Higher Dice Coefficients meant that a specific FOV normalization approach led to a higher skull stripping accuracy.

*Consistency* was measured by the standard deviation and distribution of the Dice Coefficients in the diverse images. More compact spread of Dice Coefficients, and hence smaller standard deviation, indicated that a specific FOV normalization approach led to more consistent skull stripping results.

*Success rate* was measured by the percentage of images that had a Dice Coefficient greater than a threshold. We used different threshold values, namely 0.7, 0.8, and 0.9, to define success rate — call a skull stripping result ”successful” if it has a Dice Coefficient with manual annotations over 0.7, 0.8, or 0.9.

## 4 Results

### 4.1 Overall FOV Normalization: Quantitative and Visual Results

Our first set of results was the quantitative leave-oneout cross validation in the 10 intermediate templates. The Dice Coefficients between the computed and standardized

FOVs were 0.936± 0.047. This confirmed that our automatically computed FOV and manually normalized FOV were in close agreement.

In our second set of results, we ran the FOV normalization in all 5 diverse datasets combined. Fig. 5 visualizes the normalized FOV (yellow masks on top of the original images) in 15 randomly-chosen subjects, 3 from each of the 5 datasets. Three observations could be made from Fig. 5.

One was the **effectiveness** of the proposed FOV normalization. By visual inspection, the FOV normalization was successful in all subjects. Here we defined success when the computed FOV covered the whole brain, the skull from left to right and from anterior to posterior, and ended 10–20mm inferior to the most inferior voxel in the cerebellum. Such an FOV was similar to the standard FOV in the MNI152 space we defined in Fig. 3.

The second observation was **generality**. Our framework generally applied to diverse images despite the variability in scanner, vendors, fields of strength, subject age, image resolution and orientation. And this was achieved at a fixed set of parameters and by using a fixed set of intermediate templates (all from adults).

The third observation was **robustness**. In Fig. 5, neonatal images from the BCH and ICL datasets contained shoulders and chests. The proposed FOV normalization was still successful. Two images in the ICL dataset even contained the hands of neonates in their mouths (sagittal view of the first two columns in the ICL dataset in Fig. 5). Yet, our proposed framework was also successful.

### 4.2 FOV Normalization's Effects on Skull Stripping

Visually, as Fig. 1 already showed, FOV normalization helped improve skull stripping accuracies for one subject in the ADNI dataset (elderly >70 years of age, the top row in Fig. 1) and another subject in the ICL dataset (child 2 years of age, the bottom row in Fig. 1). For the randomly chosen ADNI subject, the performances of BET, BSE and HWA improve significantly after FOV normalization. For the randomly chosen ICL subject, BET, 3dSkullStrip, BSE and HWA all performed much more accurately or even turned failure into success after FOV normalization.

Quantitatively, Fig. 6 shows the skull stripping accuracies before and after FOV normalization. For each dataset, FOV normalization significantly improved the performance of at least 3, and mostly 4, out of all 4 skull stripping algorithms/tools. In many cases, we saw the failed skull stripping results turned into reasonably good results after FOV normalization. In big data neuro-development studies (e.g., [11]), all those images need to be batch processed. The far right columns in all panels in Fig. 6 confirmed that FOV normalization significantly improved all skull stripping algorithms examined.

We note that the change in skull stripping accuracy after FOV normalization is more than marginal improvement and has meaningful impact. For many skull stripping algorithms, and for many datasets, the average Dice Coefficient increased from 0.6–0.75 to 0.85–0.9 or above. Some Dice Coefficients less than 0.8 or even less than 0.5, which meant failed skull

stripping, were significantly improved to 0.85 or above, which are reasonably good or successful skull stripping results. Such examples can be seen in BET for all the datasets, in 3dSkullStrip for the LPBA40 dataset, BSE for the LPBA40 and BCH datasets, HWA for the LPBA40 and ICL datasets, and MASS for all datasets. We will further quantify the success rate in the next subsection when comparing no FOV normalization with different FOV normalization approaches.

### 4.3 Comparison with Another FOV Truncation Tool.

Fig. 7 compares two FOV truncation/normalization approaches, in terms of their effects on skull stripping for 4 skull stripping algorithms. The far right shadowed columns in all sub-figures show that, the skull stripping accuracies were statistically higher and more consistent after our proposed FOV normalization than after FOV truncation by FSL's robustfov. And this observation applies to all 5 skull stripping algorithms being tested here. When it comes to a specific skull stripping algorithm, BET was most sensitive, BSE and HWA were least sensitive to the different FOV operations. When it comes to a specific dataset, ADNI was least influenced, meaning that different FOV normalization approaches did not make statistically significant difference for skull stripping the ADNI dataset.

Table 2 compares the success rate of skull stripping before any FOV operation, after FSL's robustfov and after our proposed FOV normalization. "Successful" skull stripping can be defined as having a Dice Coefficient greater than 0.7, 0.8 or 0.9 compared to manually annotated brain masks. At any of these thresholds, the proposed FOV normalization significantly improved the success rate when applied to these diverse datasets. And this was independent of the skull stripping algorithm used.

### 4.4 Computational Time

The proposed FOV normalization framework was implemented in UNIX-based Portable Batch System (PBS), where multiple bridged registrations ran in parallel on different computer notes of the cluster. The high-performance cluster has 127 computer nodes, and each node has 8 CPUs and 56 GB of shared virtual memory. With this configuration, the whole FOV normalization process took about 4–5 minutes per image. In the next section, we will discuss the computational load in the context of big data neuroimaging batch processing, which is our ultimate goal.

## 5 Discussion

Batch analysis of large-scale and diverse images is increasingly important in the big data era [45]. The variation of image FOVs may present a significant bottleneck problem especially for batch analysis. We demonstrated in this paper that fully-automated and tuning-free FOV normalization in diverse images is feasible, and can significantly improve performances of skull stripping for several major brain image analysis pipelines (FSL, AFNI, BrainSuite or the LONI pipeline, and FreeSurfer). The improvement was not just marginal, but statistically significant (Figs. 6 and 7). Moreover, in many cases, FOV normalization prevented skull stripping failures (Table 2).

A practical question is the general applicability of our framework. That is, whether we need to always choose intermediate templates every time we process a new dataset. We want to emphasize that our 10 intermediate templates are fixed for all testing images. There is no need to update intermediate templates. Table 1 shows that no intermediate templates were chosen from at least 3 testing datasets (1 dataset for adults and 2 datasets for 0–3 years old children), whereas the performance in those testing datasets was promising (Figures 5, 6 and 7, and Table 2). This shows the promise of generality of our framework in multi-site unseen datasets.

We defined the standard FOV in the MNI152 atlas. We could have used other atlases, such as the Talairach atlas [49,50], FreeSurfer atlases [51,52], SRI24 atlas [53], the UNC-0-1-2 years old atlas [54], and NIHPD atlases [55]. We chose the MNI152 atlas because of its wide adoption in the neuroimaging community — it is a default atlas included in many software packages, such as FSL [20], SPM [36] and AFNI [17], and many other atlases have already been spatially normalized into the MNI152 space, such as the Harvard-Oxford atlas [56], and the AAL atlas [57,58]. In defining the standard FOV in the MNI152 space, we dilated the foreground FOV mask by 5mm. This dilation is to account for possible registration errors between the MNI152 atlas and an arbitrary target image. Having this buffer zone ensures that the skull of the target image will not be unintentionally truncated. This buffer zone may result in the inclusion of some background regions outside the skull (e.g., see Fig. 5). However, this is of no concern, since such moderate background regions do not usually affect skull stripping.

Normalized FOV provides a much tractable basis for using a consistent set of parameters to skull stripping very diverse datasets. Multi-atlas skull stripping (MASS) is one of a very recently developed and widely validated algorithm. Our results in Fig. 6 show that the accuracy and consistency of MASS skull stripping results increased in two datasets involving brain MRI of adults. More importantly, the accuracy and consistency improved more significantly for pediatric brain MRI. The default atlases used in MASS are from images of adults. Therefore, Fig. 6 shows a potential of using a fixed set of adult atlases to skull strip diverse set of pediatric brain MRI. Not needing pediatric atlases opens possibilities for a single framework to be applicable to multi-site brain MRI for both adults and children. We will explore further in this direction.

There are several limitations to this study.

First, we only tested our framework in images having larger FOV than the standard FOV. There are occasions where the images are scanned intentionally with a very small FOV. For example, in brain tumor studies, multi-modal images, especially time-course dependent perfusion images, usually only scan part of the brain which contains the tumor (partial brain FOV). Our framework cannot fill the missing information into these partial brain FOV images.

A second limitation of our framework is that we have not studied how the FOV normalization accuracy and efficacy changes as the number of intermediate templates changes, or as the parameter values in the template selection (Eqn. 2) change. Independent

studies for skull stripping [33] (in their Figure 9) and for brain region segmentation [59,60,61] (in their Figures 5, 11 and 5 respectively) have shown that 8–10 atlases would saturate the segmentation accuracy. According to these studies, we empirically set our parameter values (15 intermediate templates), we succeeded in normalizing the FOV in every target image. We do plan to use this framework in larger-scale batch processing of clinical neuro-developmental and neuro-oncology images, so a more rigorous study of sensitivity to parameter settings may emerge as needed in the future.

The third issue is computational time. The proposed framework takes 4–5 minutes whereas FSL's robustfov takes about 10 seconds. Compared to robustfov, our framework in its current proof-of-concept version trades in computational time to gain increased generality, robustness, and significantly improved skull stripping performances (Figs. 5, 6 and 7). We also note that, the current batch processing of big data neuroimaging datasets is typically not done in real time. The full process includes many other steps (e.g., surface reconstruction, cortical thickness measurements, brain parcellation, etc), and skull stripping is just one early step in the processing pipeline. Adding 4–5 minutes of the proposed FOV normalization prior to skull stripping will not dramatically change the computation time of the whole pipeline, but will significantly improve the performance of the skull stripping step (Fig. 6), which is a basis for all subsequent steps. At a minimum this will reduce the failure rate for automatic processing, and further, it has the potential for improving the final outcome, a result that requires further validation. Our future work includes efforts to speed up the proposed FOV normalization, by faster pre-alignment (e.g., [62]), or by GPU speedup of the registration [63], since image registration takes more than 90% of the whole FOV normalization computational time.

The fourth issue is in datasets we have used in the evaluations. The IBSR dataset we have used was scanned back in 1996. We decided to still include this IBSR dataset among other 4 more contemporary datasets, since some valuable, large scale neuroimaging studies (e.g., Baltimore Longitudinal Study of Aging [64,65]) also began acquiring MRI data more than 20 plus years ago.

The fifth issue is that we only showed the effect of FOV normalization on skull stripping, but we did not yet study the effect of improved skull stripping on subsequent longitudinal, cross-sectional or population studies. The latter is future work. Existing studies already show that improved skull stripping can improve voxel-based morphometry (VBM) [7], can improve atrophy detection [23,66], and can improve neuro-development characterization [67,68]. Batch processing with robust skull stripping was also shown to be the basis for accurately associating genotype with image phenotype information [9,10]. Given that our proposed FOV normalization improved skull stripping performances not just marginally, but significantly (Figs. 6 and 7), and turned many failed skull stripping results into successful ones (Table 2), we believe that it will also be a useful pre-processing step to augment all aforementioned VBM, atropy, neuro-developmental and similar studies. The time, effort and cost savings due to the decrease in skull stripping failure rate can be readily appreciated. Moreover, there is also the potential for greater algorithmic accuracy in subsequent image processing steps. Further validation effort is required to make that determination.

Last, but certainly not least, we should not overstate the effect of FOV normalization. On one hand, it is true that statistical analysis revealed that FOV normalization improved the performances of all 5 skull stripping algorithms when we looked at all 5 datasets combined in our experiments (Figs. 6 and 7). It is also our observation that the proposed FOV normalization succeeded in all of our multi-site testing data (i.e., no skull or brain were removed as a result of FOV normalization, all extra/unnecessary content of shoulder or upper chest had been excluded, and all resulting normalized FOV were visually sound.) On the other hand, however, it is also true that the improvement was only marginal in some specific images, or for a specific skull stripping algorithm in a specific dataset. This can be seen by the red-cross outliers and those $p$-values greater than 0.05 in Fig. 6. Indeed, our visual inspection found several images, for which the bad skull stripping results before FOV normalization were improved only slightly after FOV normalization truncating the neck and the shoulder. This underlines the fact that FOV variation is only one factor undermining the consistency, accuracy and success rate of skull stripping. We should not expect FOV normalization alone to solve all the inconsistency, inaccuracy and failure problems in skull stripping. There are other imaging factors or patient-related anatomical factors. Dealing with those factors may need future work combining existing skull stripping algorithms or inventing new tools in the context of big data batch processing.

To summarize, the first part of this paper presented an automated FOV normalization algorithm, and showed that, without the need for tuning any parameters, the proposed FOV normalization framework can be applied to images from different sites, from different scanners, from different fields of strength, and from subjects of various ages. In the second part, we conducted a comprehensive evaluation and verified the previously untested hypothesis that FOV normalization can significantly improve skull stripping performances in general, for the four widely used skull stripping algorithms we have tested. While FOV variation is only one factor, future studies need to address other imaging or anatomical factors to further improve consistency, accuracy and success rate of skull stripping facing large-scale, diverse brain MR images, and how more consistent and successful skull stripping impacts subsequent imaging phenotype studies.

## Acknowledgments

The authors would like to acknowledge Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043 for providing support of the high-performance batch computing environment, Thrasher Research Fund Early Career Award for support to YO, Boston Children's Hospital and Harvard Medical School Faculty Development Fellowship to YO, NIH R01 EB014947 for support to YO, LZ, KR, PEG, SNM and RLG, NIH R00 HD061485 for support to LZ, NIH K23 CA169021 for support to ERG, and NIH U01 CA154601 and U24 CA180927 for support to YO, XD, ERG and JKC.

## References

1. Johansen-Berg H, Behrens TE, Sillery E, Ciccarelli O, Thompson AJ, Smith SM, and Matthews PM, "Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus," Cerebral cortex, vol. 15, no. 1, pp. 31–39, 2005. [PubMed: 15238447]
2. Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, and Cox RW, "A new method for improving functional-to-structural MRI alignment using local Pearson correlation," Neuroimage, vol. 44, no. 3, pp. 839–848, 2009. [PubMed: 18976717]

3. Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, and Toga AW, “Longitudinal mapping of cortical thickness and brain growth in normal children,” The Journal of neuroscience, vol. 24, no. 38, pp. 8223–8231, 2004. [PubMed: 15385605]

4. Bartsch AJ, Homola G, Biller A, Smith SM, Weijers H-G, Wiesbeck GA, Jenkinson M, De Stefano N, Solymosi L, and Bendszus M, “Manifestations of early brain recovery associated with abstinence from alcoholism,” Brain, vol. 130, no. 1, pp. 36–47, 2007. [PubMed: 17178742]

5. Shi F, Fan Y, Tang S, Gilmore JH, Lin W, and Shen D, “Neonatal brain image segmentation in longitudinal MRI studies,” Neuroimage, vol. 49, no. 1, pp. 391–400, 2010. [PubMed: 19660558]

6. Ou Y, Zöllei L, Retzepi K, Castro V, Bates SV, Pieper S, Andriole KP, Murphy SN, Gollub RL, and Grant PE, “Using clinically acquired mri to construct age-specific ADC atlases: Quantifying spatiotemporal adc changes from birth to 6-year old,” Human Brain Mapping, vol. 38, no. 6, pp. 3052–3068, 2017. [PubMed: 28371107]

7. Acosta-Cabronero J, Williams GB, Pereira JM, Pengas G, and Nestor PJ, “The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry,” Neuroimage, vol. 39, no. 4, pp. 1654–1665, 2008. [PubMed: 18065243]

8. Ridgway GR, Henley SM, Rohrer JD, Scahill RI, Warren JD, and Fox NC, “Ten simple rules for reporting voxel-based morphometry studies,” Neuroimage, vol. 40, no. 4, pp. 1429–1435, 2008. [PubMed: 18314353]

9. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivires S, Jahanshad N, Toro R, Wittfeld K, Abramovic L, and Andersson M, “Common genetic variants influence human subcortical brain structures,” Nature, vol. 520, pp. 224–229, 2015. [PubMed: 25607358]

10. Medland SE, Jahanshad N, Neale BM, and Thompson PM, “Whole-genome analyses of whole-brain data: working within an expanded search space,” Nature neuroscience, vol. 17, no. 6, pp. 791–800, 2014. [PubMed: 24866045]

11. Evans AC, “The NIH MRI study of normal brain development,” NeuroImage, vol. 30, pp. 184–202, Mar. 2006. [PubMed: 16376577]

12. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, L Whitwell J, and Ward C, “The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods,” Journal of Magnetic Resonance Imaging, vol. 27, no. 4, pp. 685–691, 2008. [PubMed: 18302232]

13. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, Mc-Carthy MI, Ouwehand WH, and Samani NJ, “Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls,” Nature, vol. 447, no. 7145, pp. 661–678, 2007. [PubMed: 17554300]

14. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, Olson JJ, Mikkelsen T, Lehman N, and Aldape K, “Comprehensive genomic characterization defines human glioblastoma genes and core pathways,” Nature, vol. 455, no. 7216, pp. 1061–1068, 2008. [PubMed: 18772890]

15. Smith SM, “Fast robust automated brain extraction,” Human brain mapping, vol. 17, no. 3, pp. 143–155, 2002. [PubMed: 12391568]

16. Jenkinson M, Pechaud M, and Smith S, “BET2: MR-based estimation of brain, skull and scalp surfaces,” in 11th annual meeting of the Organization for Human Brain Mapping, vol. 17, p. 167, Toronto, ON, 2005.

17. Cox RW, “AFNI: software for analysis and visualization of functional magnetic resonance neuroimages,” Computers and Biomedical research, vol. 29, no. 3, pp. 162–173, 1996. [PubMed: 8812068]

18. Shattuck DW and Leahy RM, “BrainSuite: an automated cortical surface identification tool,” Medical image analysis, vol. 6, no. 2, pp. 129–142, 2002. [PubMed: 12045000]

19. Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, and Fischl B, “A hybrid approach to the skull stripping problem in MRI,” Neuroimage, vol. 22, no. 3, pp. 1060–1075, 2004. [PubMed: 15219578]

20. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, and Smith SM, “Fsl,” Neuroimage, vol. 62, no. 2, pp. 782–790, 2012. [PubMed: 21979382]

21. Rex DE, Ma JQ, and Toga AW, "The LONI pipeline processing environment," Neuroimage, vol. 19, no. 3, pp. 1033–1048, 2003. [PubMed: 12880830]
22. Fischl B, "FreeSurfer," Neuroimage, vol. 62, no. 2, pp. 774–781, 2012. [PubMed: 22248573]
23. Popescu V, Battaglini M, Hoogstrate WS, Verfaillie SCJ, Sluimer IC, Van Schijndel RA, Van Dijk BW, Cover KS, Knol DL, and Jenkinson M, "Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3d T1 images in multiple sclerosis," Neuroimage, vol. 61, no. 4, pp. 1484–1494, 2012. [PubMed: 22484407]
24. Heckemann RA, Husson R, Ledig C, Rueckert D, and Hammers A, "Positional normalization as a first step in processing magnetic resonance brain images: work in progress," in Swedish Symposium on Image Analysis for 2013, pp. 1–3.
25. Liao M, Yang W, Zhang X, Lu M, and Dou W, "A pre-processing method for magnetic resonance images of head to improve the performance of brain extraction tools," in Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on, pp. 121–125, IEEE, 2014.
26. Iglesias JE, Liu C-Y, Thompson PM, and Tu Z, "Robust brain extraction across datasets and comparison with publicly available methods," Medical Imaging, IEEE Transactions on, vol. 30, no. 9, pp. 1617–1634, 2011.
27. Shi F, Wang L, Dai Y, Gilmore JH, Lin W, and Shen D, "LABEL: pediatric brain extraction using learning-based meta-algorithm," Neuroimage, vol. 62, no. 3, pp. 1975–1986, 2012. [PubMed: 22634859]
28. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, and Polimeni JR, "The minimal preprocessing pipelines for the Human Connectome Project," Neuroimage, vol. 80, pp. 105–124, 2013. [PubMed: 23668970]
29. Yang W, Liao M, Zhang X, Dou W, Zhang M, Chen H, Li S, Wang Y, and Dai J, "An Improvement Method of Brain Extraction Tools for Magnetic Resonance Images," Journal of Medical Imaging and Health Informatics, vol. 4, no. 6, pp. 895–900, 2014.
30. Ou Y, Akbari H, Bilello M, Da X, and Davatzikos C, "Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights," Medical Imaging, IEEE Transactions on, vol. 33, no. 10, pp. 2039–2065, 2014.
31. Mazziotta JC, Toga AW, Evans A, Fox P, and Lancaster J, "A probabilistic atlas of the human brain: theory and rationale for its development the international consortium for brain mapping (ICBM)," Neuroimage, vol. 2, no. 2PA, pp. 89–101, 1995. [PubMed: 9343592]
32. Ou Y, Sotiras A, Paragios N, and Davatzikos C, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," Medical image analysis, vol. 15, no. 4, pp. 622–639, 2011. [PubMed: 20688559]
33. Doshi J, Erus G, Ou Y, Gaonkar B, and Davatzikos C, "Multi-Atlas Skull-Stripping," Academic radiology, vol. 20, no. 12, pp. 1566–1576, 2013. [PubMed: 24200484]
34. Lee J-M, Yoon U, Nam SH, Kim J-H, Kim I-Y, and Kim SI, "Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error," Computers in biology and medicine, vol. 33, no. 6, pp. 495–507, 2003. [PubMed: 12878233]
35. Fennema-Notestine C, Burak Ozyurt I, Clark CP, Morris S, Bischoff-Grethe A, Bondi MW, Jernigan TL, Fischl B, Segonne F, and Shattuck DW, "Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location," Human brain mapping, vol. 27, no. 2, pp. 99–113, 2006. [PubMed: 15986433]
36. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, and Nichols TE, Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images. Academic press, 2011.
37. Evans AC, Janke AL, Collins DL, and Bail-let S, "Brain templates and atlases," Neuroimage, vol. 62, no. 2, pp. 911–922, 2012. [PubMed: 22248580]
38. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, and Buckner RL, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," Journal of cognitive neuroscience, vol. 19, no. 9, pp. 1498–1507, 2007. [PubMed: 17714011]

39. Shattuck DW, Prasad G, Mirza M, Narr KL, and Toga AW, “Online resource for validation of brain segmentation methods,” NeuroImage, vol. 45, no. 2, pp. 431–439, 2009. [PubMed: 19073267]
40. Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, and Toga AW, “Construction of a 3d probabilistic atlas of human cortical structures,” Neuroimage, vol. 39, no. 3, pp. 1064–1080, 2008. [PubMed: 18037310]
41. Warfield SK, Zou KH, and Wells WM, “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation,” Medical Imaging, IEEE Transactions on, vol. 23, no. 7, pp. 903–921, 2004.
42. Worth A, “The internet brain segmentation repository (ibsr),” 1996.
43. Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, Edwards AD, and Hammers A, “Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest,” Neuroimage, vol. 40, no. 2, pp. 672–684, 2008. [PubMed: 18234511]
44. Leung KK, Barnes J, Modat M, Ridgway GR, Bartlett JW, Fox NC, Ourselin S, and Initiative ADN, “Brain MAPS: an automated, accurate and robust brain extraction technique using a template library,” Neuroimage, vol. 55, no. 3, pp. 1091–1108, 2011. [PubMed: 21195780]
45. Wang Y, Nie J, Yap P-T, Li G, Shi F, Geng X, Guo L, Shen D, and for the Alzheimer’s Disease Neuroimaging Initiative, “Knowledge-Guided Robust MRI Brain Extraction for Diverse Large-Scale Neuroimaging Studies on Humans and Non-Human Primates,” PLoS ONE, vol. 9, p. e77810, Jan. 2014. [PubMed: 24489639]
46. Hartley S, Scher A, Korf E, White L, and Launer L, “Analysis and validation of automated skull stripping tools: A validation study based on 296 MR images from the Honolulu Asia aging study,” NeuroImage, vol. 30, pp. 1179–1186, 5 2006. [PubMed: 16376107]
47. Galdames FJ, Jaillet F, and Perez CA, “An accurate skull stripping method based on simplex meshes and histogram analysis for magnetic resonance images,” Journal of neuroscience methods, vol. 206, no. 2, pp. 103–119, 2012. [PubMed: 22387261]
48. Sadananthan SA, Zheng W, Chee MW, and Zagorodnov V, “Skull stripping using graph cuts,” NeuroImage, vol. 49, no. 1, pp. 225–239, 2010. [PubMed: 19732839]
49. Lancaster JL, Summerlin JL, Rainey L, Freitas CS, and Fox PT, “The Talairach Daemon, a database server for Talairach atlas labels,” Neuroimage, vol. 5, no. 4, p. S633, 1997.
50. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, and Fox PT, “Automated Talairach atlas labels for functional brain mapping,” Human brain mapping, vol. 10, no. 3, pp. 120–131, 2000. [PubMed: 10912591]
51. Desikan RS, Sgonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, and Hyman BT, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” Neuroimage, vol. 31, no. 3, pp. 968–980, 2006. [PubMed: 16530430]
52. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Sgonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, and Kennedy D, “Automatically parcellating the human cerebral cortex,” Cerebral cortex, vol. 14, no. 1, pp. 11–22, 2004. [PubMed: 14654453]
53. Rohlfing T, Zahr NM, Sullivan EV, and Pfefferbaum A, “The SRI24 multichannel atlas of normal adult human brain structure,” Human brain mapping, vol. 31, no. 5, pp. 798–819, 2010. [PubMed: 20017133]
54. Shi F, Yap P-T, Wu G, Jia H, Gilmore JH, Lin W, and Shen D, “Infant brain atlases from neonates to 1-and 2-year-olds,” PLoS One, vol. 6, no. 4, p. e18746, 2011. [PubMed: 21533194]
55. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, and Group BDC, “Unbiased average age-appropriate atlases for pediatric studies,” NeuroImage, vol. 54, no. 1, pp. 313–327, 2011. [PubMed: 20656036]
56. Caviness V, Meyer J, Makris N, and Kennedy D, “MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability,” Cognitive Neuroscience, Journal of, vol. 8, no. 6, pp. 566–587, 1996.
57. Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, and Petrides M, “Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space,” Neuroimage, vol. 10, no. 3, pp. 233–260, 1999. [PubMed: 10458940]

58. Schmahmann JD, Doyon J, Petrides M, Evans AC, and Toga AW, MRI atlas of the human cerebellum. Academic Press, 2000.

59. Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D, and Initiative ADN, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," Neuroimage, vol. 49, no. 3, pp. 2352–2365, 2010. [PubMed: 19857578]

60. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, and Rueckert D, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," Neuroimage, vol. 46, no. 3, pp. 726–738, 2009. [PubMed: 19245840]

61. Doshi J, Erus G, Ou Y, Resnick SM, Gur RC, Gur RE, Satterthwaite TD, Furth S, Davatzikos C, Initiative AN, and others, "MUSE: MUlti-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection," NeuroImage, 2016.

62. Ardekani BA and Bachman AH, "Model-based automatic detection of the anterior and posterior commissures on MRI scans," Neuroimage, vol. 46, no. 3, pp. 677–682, 2009. [PubMed: 19264138]

63. Shams R, Sadeghi P, Kennedy R, and Hartley R, "A survey of medical image registration on multicore and the GPU," Signal Processing Magazine, IEEE, vol. 27, no. 2, pp. 50–60, 2010.

64. Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, and D'agostino RB, "Lifetime risk of dementia and Alzheimer's disease The impact of mortality on risk estimates in the Framingham Study," Neurology, vol. 49, no. 6, pp. 1498–1504, 1997. [PubMed: 9409336]

65. Pacheco J, Goh JO, Kraut MA, Ferrucci L, and Resnick SM, "Greater cortical thinning in normal older adults predicts later cognitive impairment," Neurobiology of aging, vol. 36, no. 2, pp. 903–908, 2015. [PubMed: 25311277]

66. Battaglini M, Smith SM, Brogi S, and De Stefano N, "Enhanced brain extraction improves the accuracy of brain atrophy estimation," Neuroimage, vol. 40, no. 2, pp. 583–589, 2008. [PubMed: 18255315]

67. Erus G, Battapady H, Satterthwaite TD, Hakonarson H, Gur RE, Davatzikos C, and Gur RC, "Imaging patterns of brain development and their relationship to cognition," Cerebral Cortex, p. bht425, 2014.

68. Ou Y, Gollub RL, Retzepi K, Reynold NA, Pienaar R, Murphy SN, Grant PE, and Zllei L, "Brain Extraction in Pediatric ADC Maps, toward Characterizing Neuro-Development in Multi-Platform and Multi-Institution Clinical Images," NeuroImage, vol. 122, pp. 246–261, 2015. [PubMed: 26260429]

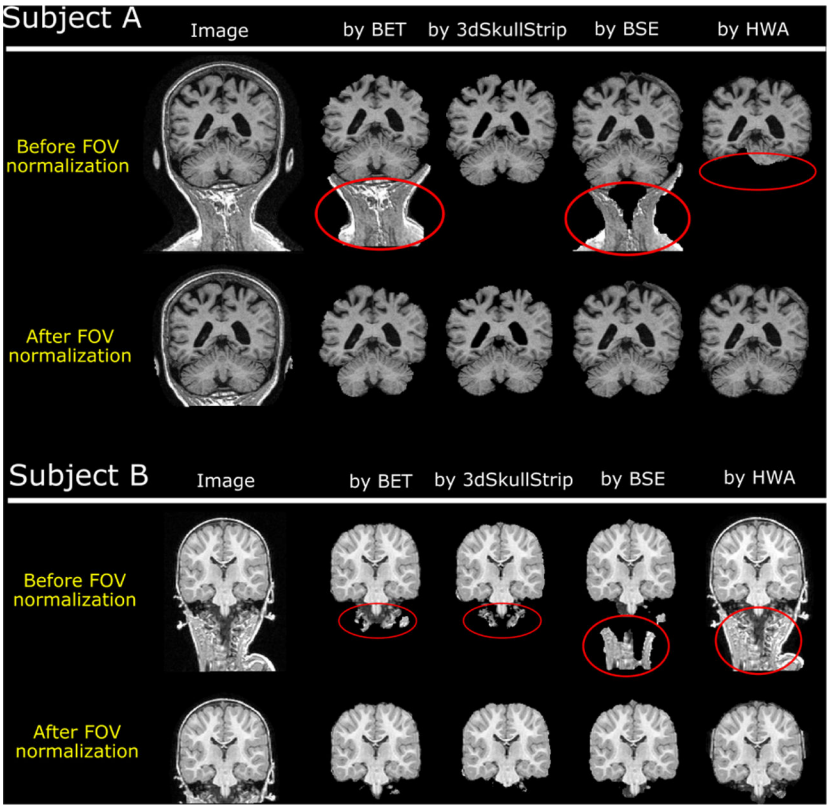


**Fig. 1.**

FOV variation is one factor contributing to inconsistency, inaccuracy and failure in skull stripping. Subject A is a 62 year old patient from the public ADNI dataset. Subject B is a 2 year old patient from the ICL public dataset. Red contours highlight errors that are present in images with the original FOVs but are largely gone in images after FOV normalization.

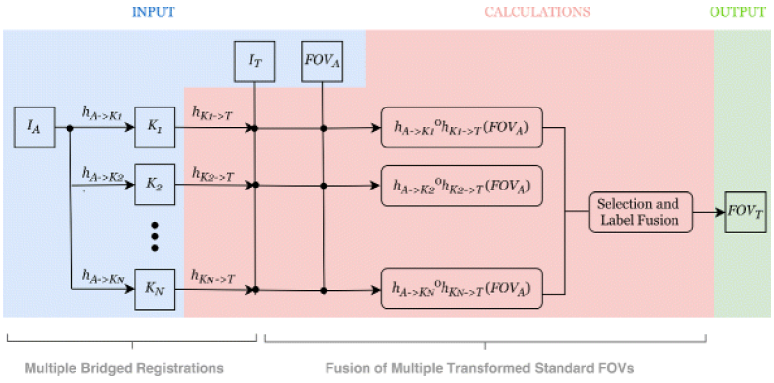


**Fig. 2.**
Diagram of the proposed normalizeFOV framework.

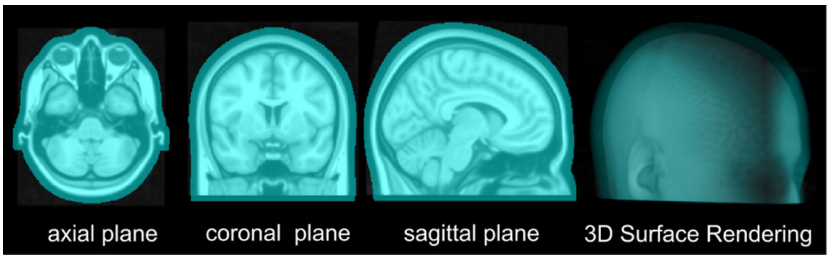


**Fig. 3.**

The definition of the standard FOV used in this paper (the blue mask), overlaid on the MNI152 atlas.

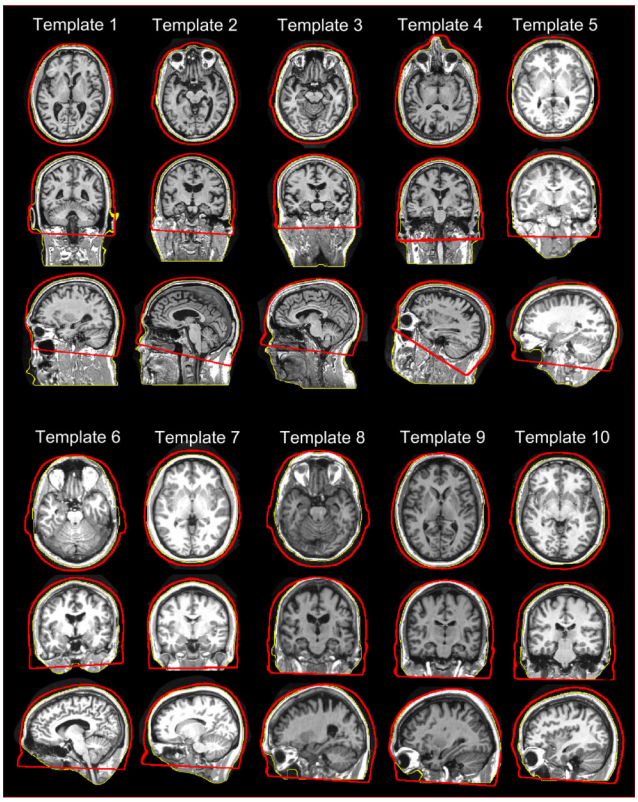


**Fig. 4.**

The intermediate templates. Their original FOVs are outlined by the yellow contours. The red contours outline the standard FOV warped from the MNI152 space to the template space, i.e., $h_{A \to K_n} \circ h_{K_n \to T}(FOV_A)(n = 1, \ldots, N)$ when $K_n = T\left(\text{i.e.}, h_{K_n \to T} = \textbf{Identity}\right)$.

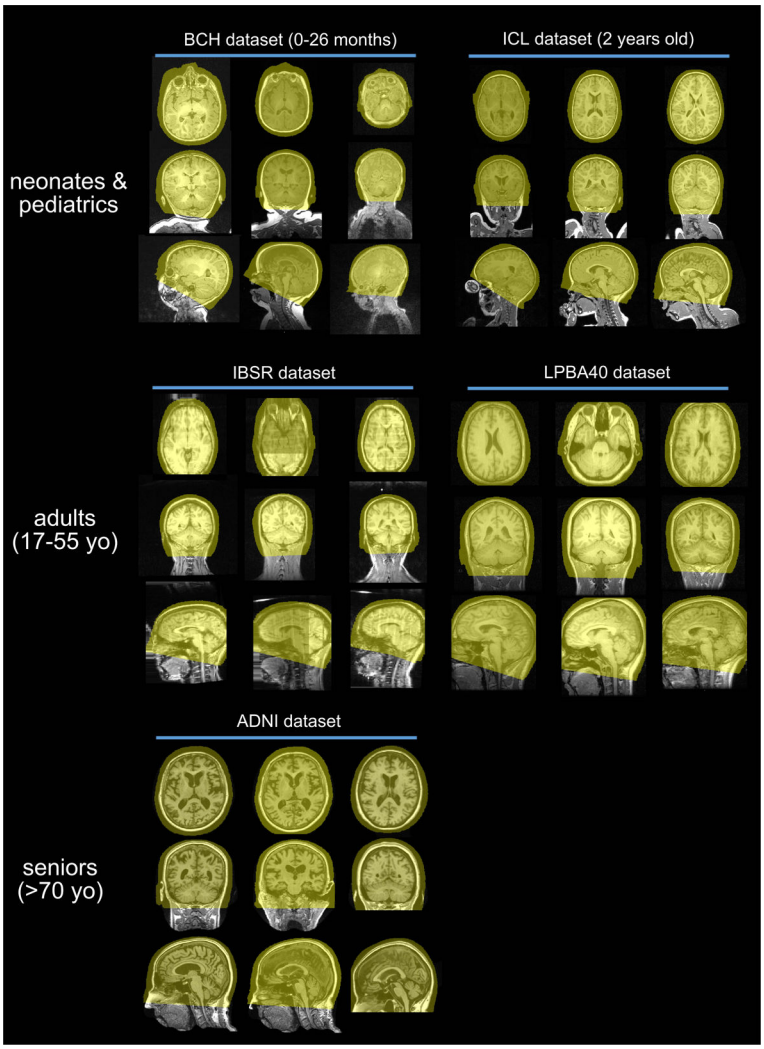


**Fig. 5.**

Normalized FOVs (yellow masks) overlaid on 15 images from subjects of various ages. For each of the 5 datasets, there are images from 3 subjects displayed in 3 columns, each column containing axial, coronal and sagittal views.

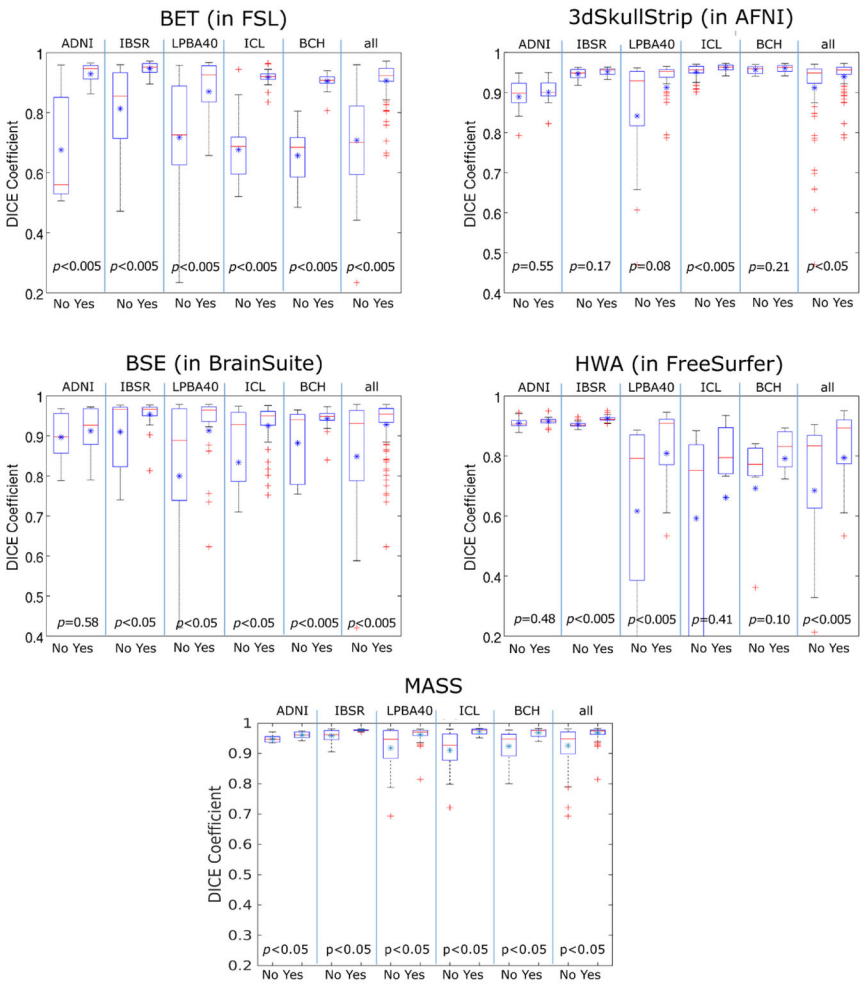


**Fig. 6.**

Skull stripping accuracy in diverse datasets with (yes) or without (no) FOV normalization, for 4 different skull stripping algorithms/tools. The asterisks in the box and whisker plots are the mean values. *p*-values are provided for the test of statistical significance.

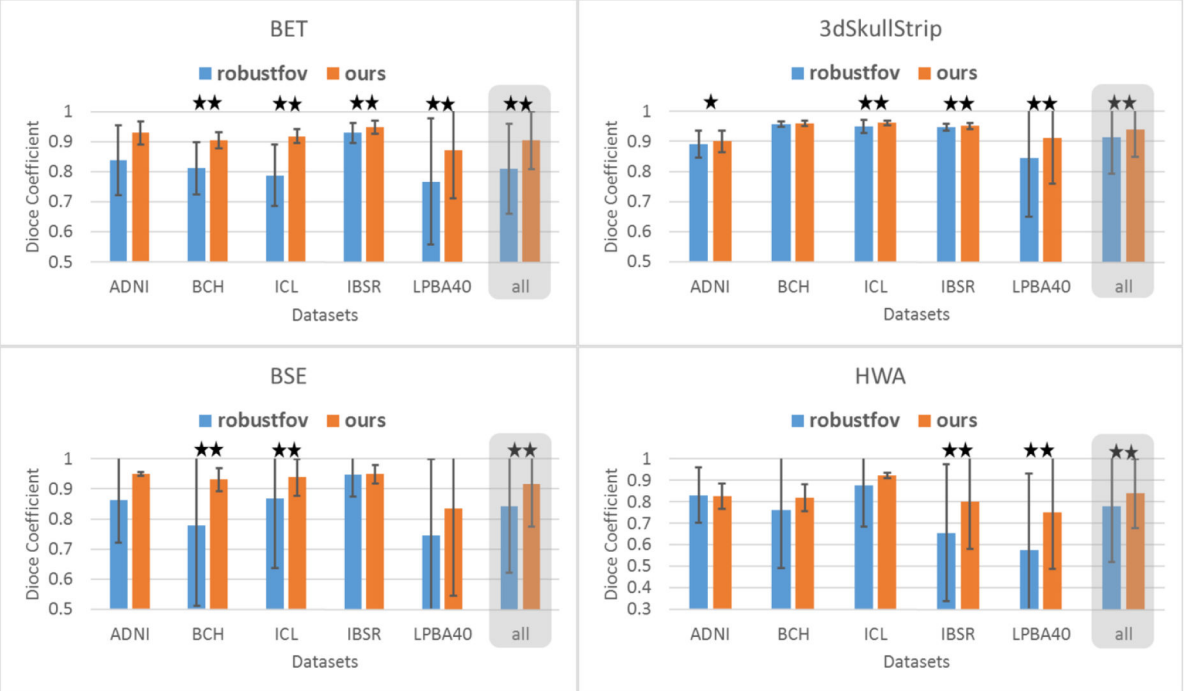


**Fig. 7.**

Comparison of FSL's robustfov and our proposed FOV normalization framework, in terms of their effects on the performances of 4 skull stripping algorithms in 5 diverse datasets. One ⋆ means $p$-value < 0.05 and two ⋆'s means $p$-value < 0.01.

**Table 1**

Information about the 5 datasets used in the paper to test our algorithm. The last row shows that no intermediate templates were chosen from 3 testing datasets, including one adult dataset namely IBSR as well as two young children datasets namely ICL and BCH. Indeed, 3 other intermediate templates were chosen from the OASIS dataset which are different from any of the 5 testing datasets listed below. ADNI – Alzheimer's Disease Neuroimaging Initiate; IBSR – Internet Brain Segmentation Repository; LPBA – LONI Probabilistic Brain Atlas; MGH – Massachusetts General Hospital; UCLA – University of California at Los Angeles; ICL – Imperial College London; BCH – Boston Childrens Hospital; SPGR - Spoiled Gradient Echo. CSE-FS - conventional SE fat suppressed. EP - Echo Planar. Yrs - years. Mons - months. The BCH dataset was proprietary with publication pending but all others were public datasets.

| | **ADNI [12]** | **IBSR [42]** | **LPBA40 [40]** | **ICL [43]** | **BCH** |
|---|---|---|---|---|---|
| #Subjects | 10 | 20 | 40 | 33 | 23 |
| Age range | 55–90 yrs | 20–38 yrs | 19–40 yrs | 1.5–3 yrs (21–34 mons) | 0–2.2 yrs (0–26 mons) |
| Site | Multi-site | MGH | UCLA | ICL | BCH |
| Scanner | Siemens 3T, GE 3T, 1.5T | Siemens 1.5T, GE 1.5T | GE 1.5T | Philips 1.0T | Siemens 3T |
| Sequence | MP-RAGE SPGR | SPGR | SPGR | CSE-FS | EP |
| Image size | 200 × 256 × 200 | 256 × 256 × 128 | 256 × 256 × 128 | 156 × 256 × 256 | 200 × 200 × 176 |
| Voxel size | $1 \times 1 \times 1mm$ | $1 \times 1 \times 1.5mm$ | $1 \times 1 \times 1.5mm$ | $1 \times 1 \times 1mm$ | $1 \times 1 \times 1mm$ |
| # intermediate templates from this dataset | 4 | | 3 | | |

*Neuroinformatics*. Author manuscript; available in PMC 2020 July 04.

**Table 2**

How different FOV operations (no operation, FSL's robustfov and ours) impact the success rates of skull stripping. To draw a more general observation, we used 4 skull stripping algorithms, and we define "success" by different thresholds of Dice overlap with manually annotated brain masks. The observation is that, regardless of which skull stripping algorithm we used, regardless of how to define skull stripping "success", our proposed FOV normalization significantly improved success rates compared to FSL's robustfov, and especially compared to no FOV operations.

| | **BET** | | |
|---|---|---|---|
| | **no FOV operation** | **after FSL robustfov** | **after our FOV normalization** |
| Dice>0.7 | 51% | 83% | **98%** |
| Dice>0.8 | 30% | 62% | **95%** |
| Dice>0.9 | 16% | 37% | **78%** |

| | **3dSkullStrip** | | |
|---|---|---|---|
| | **no FOV operation** | **after FSL robustfov** | **after our FOV normalization** |
| Dice>0.7 | 94% | 96% | **100%** |
| Dice>0.8 | 91% | 91% | **98%** |
| Dice>0.9 | 83% | 84% | **91%** |

| | **BSE** | | |
|---|---|---|---|
| | **no FOV operation** | **after FSL robustfov** | **after our FOV normalization** |
| Dice>0.7 | 72% | 83% | **97%** |
| Dice>0.8 | 61% | 69% | **94%** |
| Dice>0.9 | 54% | 66% | **84%** |

| | **HWA** | | |
|---|---|---|---|
| | **no FOV operation** | **after FSL robustfov** | **after our FOV normalization** |
| Dice>0.7 | 75% | 82% | **94%** |
| Dice>0.8 | 65% | 79% | **81%** |
| Dice>0.9 | 41% | 42% | **52%** |