A Generative Skull Stripping of Multiparametric Brain MRIs Using 3D Convolutional Neural Network

Linmin Pei
Frederick National Laboratory for Cancer Research
AK Murat
University of Pittsburgh
Nourel Tahon
University of Pittsburgh
Serafettin Zenkin
University of Pittsburgh
Safa Karawi
University of Pittsburgh
Abdallah Kamal
University of Pittsburgh
Mahir Yilmaz
University of Pittsburgh
Lingling Chen
University of Pittsburgh
Mehmet Er
University of Pittsburgh
Rivka Colen (colenrr@upmc.edu)
University of Pittsburgh

Research Article

Keywords: Skull stripping, brain extraction, glioblastoma, 3D convolutional neural network, multiparametric MRIs

Posted Date: January 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1251147/v1

A Generative Skull Stripping of Multiparametric Brain MRIs Using 3D Convolutional Neural Network

Linmin Pei[1], AK Murat[2,3], Nourel Hoda M Tahon[2,3], Serafettin Zenkin[2,3], Safa Karawi[2,3], Abdallah Kamal[2,3], Mahir Yilmaz[2,3], Lingling Chen[2,3], Mehmet Er[2,3], Rivka Colen[2,3,*]

Linmin.pei@nih.gov, akm@upmc.edu, NMTahon@mdanderson.org, zenkins@upmc.edu, alkarawi.safa@gmail.com, {kamala, yilmazm2, linglingc, erme, colenrr}@upmc.edu

1, Imaging and Visualization Group, ABCS, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA.

2, Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15213, USA

3, Hillman Cancer Center, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

*, Corresponding author

Abstract-Accurate skull stripping helps following neuro-image analysis. For computer-aided methods, the presentence of the brain skull in structural MRI impacts brain tissue identification, which could result in serious misjudgment, especially for patients with brain tumors. Though there are some existing works on skull stripping in literature, most of them either focus on healthy brain MRI or only apply for a single image modality. These methods may be not optimal for multiparametric MRI scans. In the paper, we propose an ensemble neural network (EnNet), a 3D convolutional neural network (3DCNN) based method, for brain extraction of multiparametric brain MRI scans. We comprehensively investigate the skull stripping performance by using the proposed method on a total of 15 image modality combinations. The comparison shows that using all modalities provides the best performance on skull stripping. We have collected a retrospective dataset of 815 cases with glioblastoma at the University of Pittsburgh Medical Center (UPMC) and The Cancer Imaging Archive (TCIA). The ground truths of the skull stripping are verified by at least one qualified radiologist. The quantitative evaluation gives an average dice score coefficient and Hausdorff distance at 95th percentile, respectively. We also compare the performance to the state-of-the-art methods/tools. The proposed method offers the best performance.

The contributions of the work have five folds: First, the proposed method is a fully automatic end-to-end for skull stripping using a 3D deep learning method. Second, it is applicable for multiparametric MRI (mpMRIs) and is also easy to customize for a single MRI modality. Third, the proposed method not only works for healthy brain mpMRIs but also pre-/post-operative brain mpMRIs with GBM. Fourth, the proposed method is capable to handle multicenter data. Last, to the best of our knowledge, we are the first group to quantitatively compare the skull stripping performance using different modalities.

Keywords-Skull stripping, brain extraction, glioblastoma, 3D convolutional neural network, multiparametric MRIs

I. Introduction

In the U.S., there are about 23 per 100,000 population diagnosed with brain tumors during 2011-2015 [1]. Gliomas, originate from glial cells, are the most common primary brain malignancies, with varying degrees of aggressiveness [2]. To make proper treatment planning, accurate brain tumor detection and segmentation are strongly demanding. Due to time-consuming, inter-rater prone error, and low efficacy, manual brain tumor segmentation by radiologists is very challenging, and not feasible for large-scale data [3]. Therefore, automatically computer-aided brain tumor segmentation/detection is highly desired [3-9]. However, the high-resolution brain magnetic resonance image (MRI) contains some non-brain tissues, such as eyeball, skin, neck, skin, and muscle [10]. The presentence of the non-brain tissues is one of the major challenges for automatic brain image analysis. The non-brain tissues removal is a typical preprocessing step for most brain MRI studies, e.g., brain volumetric measurement [11], brain tissue segmentation [12], assessing schizophrenia [13], and Alzheimer's disease [14]. Consequently, before applying for an automatic computational technique for brain MRI studies, skull stripping is a prerequisite for brain imaging analysis [15].

As a preprocessing step, skull stripping, aka brain extraction, is to remove the skull and other non-brain tissues out from the MRI scans. It reduces human rater variance and eliminate time-consuming manual processing steps that

potentially impede not only the analysis but also the reproducibility of large-scale studies [16]. The quality of skull stripping can be affected by several reasons, including imaging artifacts, MRI scanner, and acquisition protocol, etc. Furthermore, variability of anatomy, age, and the extent of brain atrophy, has impact on skull stripping as well [17]. The problem becomes more complicated when considering MRI scans with pathological conditions, such as brain tumors. Brain tumors change the presentence of the intensity in MRI. The situation could become worse when dealing with post-treatment of the MRI with brain tumors, especially with resection surgery. The cavities resulting from resection not only change the reflection of intensity but also alters the brain anatomy. All these factors above undermine the performance of skull stripping.

We argue that a good skull stripping leads to a good following-up brain analysis. Therefore, in the paper, we propose a 3D deep neural network-based method for skull stripping. The contribution of this work includes: first, it is a fully automatic end-to-end technique for skull stripping using a 3D deep learning method; second, it is applicable for multiparametric MRI (mpMRIs) and is also easy to customize for a single MRI modality; third, it works not only for healthy brain MRI, but also for pre-/post-operative brain MRI with a brain tumor; third, the proposed method applies to multicenter data; fourth, as the best of our knowledge, we are the first group to quantitatively compare the skull stripping performance using different modalities.

II. Previous Work

There are lots of skull stripping methods proposed in literature. These methods can be broadly classified into four categories: morphology-based, intensity-based, deformable surface-based, and atlas-based [10]. The morphology-based methods utilize a morphological erosion and dilation operations to remove skulls from the brain. Brummer *et al.* propose an automatic skull stripping on MRI using a morphology-based method [18]. It combines histogram-based thresholding and morphological operations for skull stripping. Similar work proposed in [19], authors perform a 2D Marr-Hildreh operator to achieve edge detection, then employ several morphological operations for skull stripping. However, it is difficult to find the optimal morphology-based method. In addition, the proposed methods are sensitive to small data variations. Proper thresholding and edge detection are the challenges for these methods. For intensity-based methods, they separate the brain and non-brain according to the image intensity. A typical technique of the method is a watershed algorithm. The watershed algorithm extracts foreground and background, and then using markers to make watershed run and detect the exact boundaries. Hahn *et al*. utilize the watershed algorithm to remove skull on T1-weighted MR images [20]. There are some similar works, such as [21, 22]. These methods depend on the correctness of intensity distribution modeling and are sensitive to intensity bias. The deformable surface-based methods evolve and deform an active contour to fit the brain surface. A popular tool named the Brain extraction tool (BET) employs a deformable model for separating brain and non-brain from MRI [23]. BET2 is the extension of BET, which generates a better result based on a pair of T1- and T2-weighted MRI [24]. Other work, such as [25, 26] also use the deformable surface-based method for the skull stripping. However, these methods rely on the location of the initial curve and the image gradient [10]. The atlas-based methods use the transferring knowledge of the anatomical structure of a template to separate skull and brain, such as work [27, 28]. However, these atlas-based methods highly rely on the quality of image registration. Moreover, these methods are not applicable for the case with brain tumors/diseases.

In recent years, because of computer hardware development and big data availability, deep learning has been becoming prevalent in many domains, such as image analysis [29, 30], natural language processing (NLP) [31], computer vision [32], speech recognition [33], etc. Deep learning-based methods are also applied to medical image analysis, including brain segmentation [34], brain tumor classification [35], brain tumor segmentation [7], and lung cancer segmentation [36], etc. The deep learning-based methods also apply for skull stripping, such as [37-39]. However, these works may be only applicable for normal healthy brain skull stripping, or pre-operative brain with gliomas. Therefore, to overcome the limitations mentioned above, we propose a 3D convolutional neural network (3DCNN)-based end-to-end method for a generative skull stripping. It not only works for healthy brain MRIs, but also for pre-/post-operative brain MRIs with GBM. Furthermore, it is applicable for multicenter data.

III. Results

In the section, we first share the overall performance of skull stripping using the proposed method, then investigate the performance difference for several conditional MRIs (healthy brain MRIs, pre-operative brain MRIs, and post-

operative brain MRIs), subsequentially estimate the model robustness across multicenter data, and finally compare with state-of-the-arts.

a. Overall Performance of Skull Stripping

As of the combination of all image sequences provides the best performance, we employ the best model for the testing data in the testing phase. With the total number of 216 testing cases, our algorithm offers an average dice of 0.9851 ± 0.017. The complete evaluation metrics are shown in Table 1.

Table 1. Overall performance of skull stripping in testing phase.

	Dice	Precision	Recall	FPR	FNR	HD95 (mm)
Ave	0.9850	0.9940	0.9768	0.0012	0.0232	2.6098
Std	0.0171	0.0093	0.0307	0.0019	0.0307	2.4814

b. Generality of the Model

As discussed early, the proposed method works not only in healthy brain MRIs, but also in pre-/post- operative MRIs. To quantitatively evaluate the performance difference, we set up an experiment. The result is showing in Table 2. An interesting thing we noticed is that the best results happen in pre-operative brain tumor MRIs, rather than in healthy brain MRIs. The reason may be that the training data of the model are from the pre-operative mpMRIs with glioblastoma. Overall, the skull stripping performance is stable in all conditions, either the healthy brain MRIs, or brain tumor MRIs. There are 3 showcases shown in Figure 1.

Table 2. Performance comparison of skull stripping for different stage MRIs of UPMC data. The best result is highlighted in **bold**.

Type of MRI	# of cases	Dice	Precision	Recall	FPR	FNR	HD95 (mm)
Healthy brain MRIs	57	0.9851 $\pm$ 0.0139	**0.9963 $\pm$ 0.0070**	0.9745 $\pm$ 0.0260	**0.0007 $\pm$ 0.0013**	0.0255 $\pm$ 0.026	2.4399 $\pm$ 1.568
Pre-operative brain tumor MRIs	57	**0.9906 $\pm$ 0.0068**	0.9910 $\pm$ 0.0111	**0.9904 $\pm$ 0.0097**	0.0019 $\pm$ 0.0027	**0.0096 $\pm$ 0.0097**	**2.1655 $\pm$ 1.5278**
Post-operative brain tumor MRIs	63	0.9894 $\pm$ 0.0213	0.9942 $\pm$ 0.0080	0.9852 $\pm$ 0.0358	0.0011 $\pm$ 0.0016	0.0148 $\pm$ 0.0358	2.1751 $\pm$ 3.6873

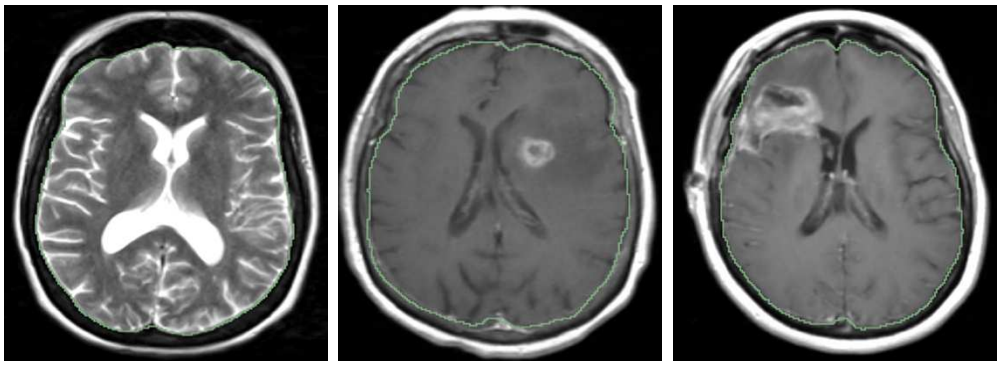

Figure 1. Showcases of skull stripping in different stage: healthy brain T2-weighted MRI (left), pre-operative T1-ce brain tumor MRI (middle), and post-operative T1-ce brain tumor MRI (right). The green contour is the boundary of skull stripping using the proposed method.

c. Model Robustness across Multicenter

It is common that brain MRIs are acquired from multiple centers/institutes using different acquisition machines or following different protocols. The multicenter issue may undermine the performance of a model training with a single-center data. In this work, we also investigate the model robustness across multicenter. Additional to our in-house UPMC data (177 cases), we randomly take 39 cases (20 pre-operative cases and 19 post-operative cases) from TCIA that collects MRIs datasets from multiple institutes/hospitals. The experimental result is summarized in Table 3.

Table 3. Skull stripping performance across multicenter (UPMC and TCIA). The best result is highlighted in **bold**.

Center	# of cases	Dice	Precision	Recall	FPR	FNR	HD95 (mm)
UPMC	177	**0.9884 ± 0.0155**	0.9939 **± 0.0091**	**0.9834 ± 0.0272**	0.0012 ± 0.002	**0.0166** ± 0.072	**2.2573** ± 2.516
TCIA	39	0.9699 ± 0.0016	**0.9946** ± 0.0105	0.9409 ± 0.0281	**0.0010 ± 0.0018**	0.0531 **± 0.0281**	4.2099 **± 1.52**

The comparison of the summary indicates that the performance at TCIA is around 2% lower than that of data obtaining from the same center for model training. However, the skull stripping performance across multicenter achieves good enough for following medical image analysis.

d. Comparison of State-Of-The-Art

In the work, we also compare the performance of skull stripping using the proposed deep learning-based method to the popular methods/tools. In doing so, we either re-implement the algorithm, or directly use the published tools. The popular methods/tools include Brain Extraction Tool (BET) [23], 3d skull stripping (3dSS) [42], Robust Learning-Based Brain Extraction (ROBEX) [43], UNet 3D (UNet3D) [37], and DeepMedic by UPNN [38]. The first three tools are using traditional machine learning-based methods, and the last two are using deep learning-based methods. An example case showing contours overlaid with the multiparametric sequence is showing in Figure 2. The performance comparison is showing in Figure 3 and Table 4.

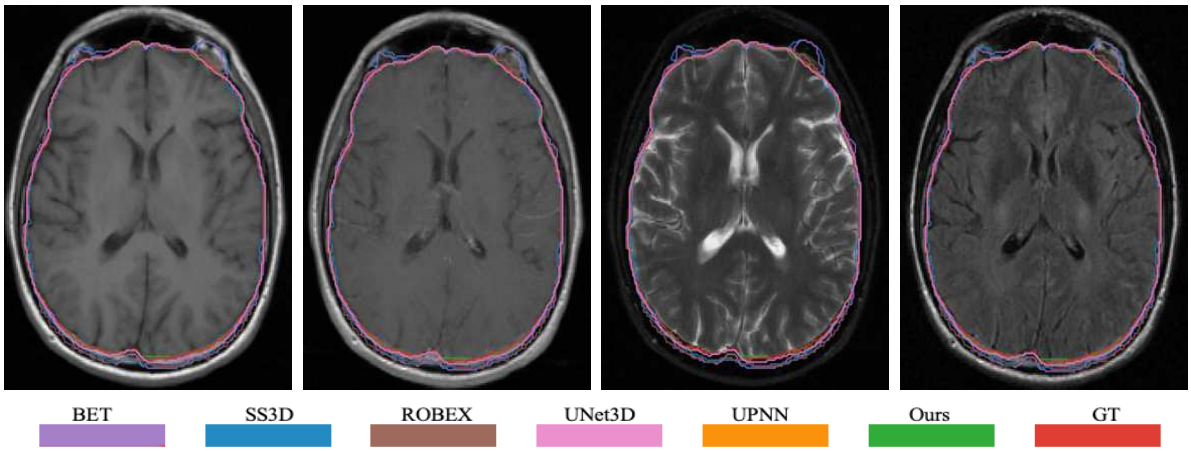

Figure 2. An example of skull stripping contours in color overlaid with T1, T1ce, T2, and T2-FLAIR (from left to right) using different methods/tools (color image for a better visualization). GT is the ground truth.

Table 4. Skull stripping performance comparison to state-of-the-arts. The best result is highlighted in bold.

	# of cases	Dice	Precision	Recall	FPR	FNR	HD95 (mm)
BET [23]	216	0.8494 ± 0.0455	0.7463 ± 0.0718	0.9916 **± 0.0186**	0.0650 ± 0.0224	0.0084 **± 0.0186**	19.9951 ± 5.4941
3dSS [42]	216	0.8427 ± 0.0449	0.7430 ± 0.0751	0.9809 ± 0.0279	0.0660 ± 0.0238	0.0191 ± 0.0279	19.9087 ± 4.4316
ROBEX [43]	216	0.9555 ± 0.0173	0.9730 ± 0.0236	0.9396 ± 0.0318	0.0053 ± 0.0057	0.0604 ± 0.0318	4.4792 **± 1.8869**
UNet3D [37]	216	0.9773 ± 0.0179	0.9818 ± 0.0168	0.9735 ± 0.0290	0.0035 ± 0.0034	0.0265 ± 0.0290	3.1219 ± 2.7262
UPNN [38]	216	0.9743 ± 0.0257	0.9814 ± 0.0156	0.9684 ± 0.0405	0.0035 ± 0.0032	0.0316 ± 0.0405	3.3924 ± 2.8434
EnNet (ours)	216	**0.9850 ± 0.0171**	**0.9940 ± 0.0093**	**0.9768** ± 0.0307	**0.0012 ± 0.0019**	**0.0232** ± 0.0307	**2.6098** ± 2.4814

The performance comparison demonstrates that the proposed method offers the best results in terms of the dice, precision, recall, FPR, FNR, and the HD95. The small value of the standard deviation indicates the robustness of the skull stripping performance.

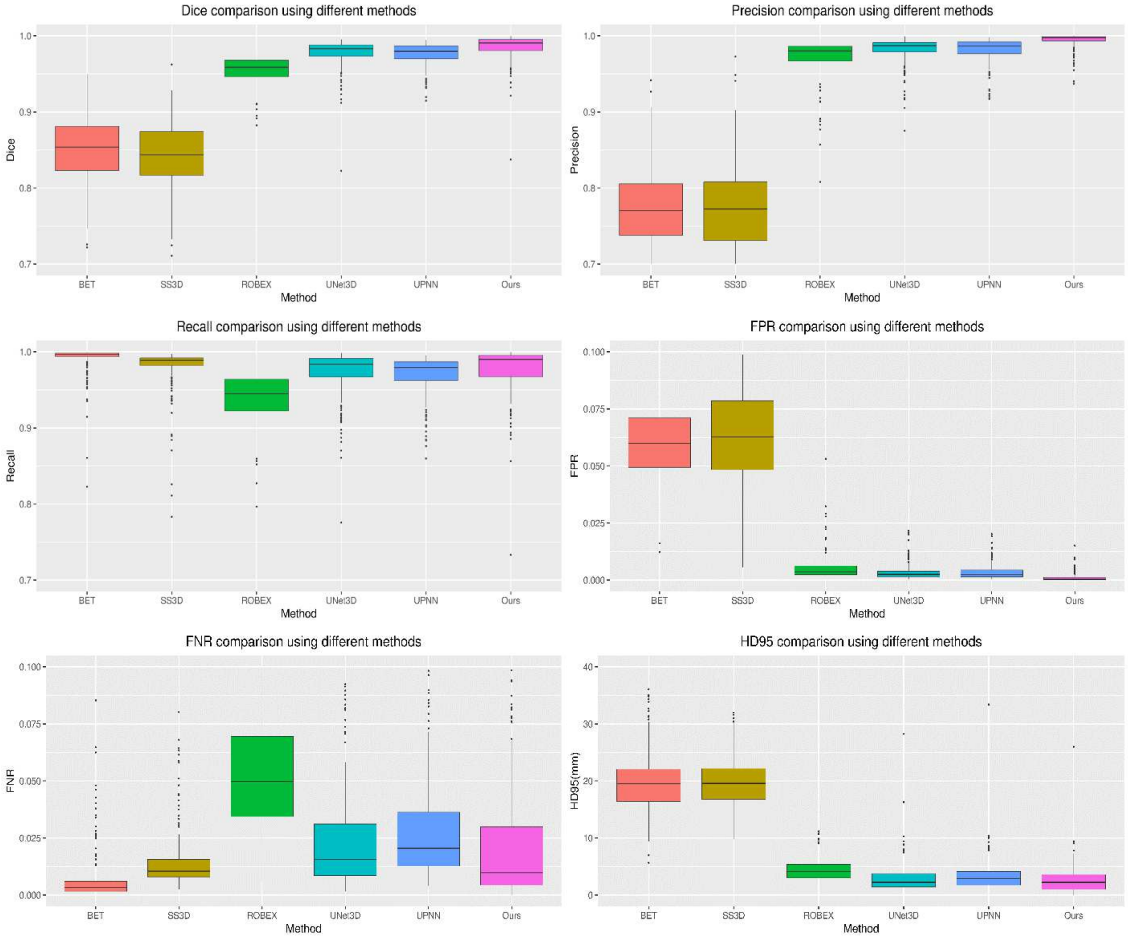

Figure 3. Box plot of performance comparison to state-of-the-arts on dice (top left), precision (top right), recall (middle left), FPR (middle right), FNR (bottom left), and HD95 (bottom right), respectively.

IV. Discussion

Even though there are extensive works on skull stripping in literature [16, 24, 37-39], to best of our knowledge, none of the methods/algorithms have explicitly quantitative analysis of performance on different image sequence combinations. It is known that different image provides different brain information, therefore, multiparametric MRIs are widely used in radiomics brain research, include brain segmentation, and brain tumor segmentation. In this work, we are the first group quantitatively showing the performance difference with different image sequence combinations. In the training phase, we randomly take 480 cases as the training dataset, and 119 cases as the validation dataset. We take the hyper-parameter setting as discussed in Section VI. The dice and loss change in the training phase and in the validation phase are plotted in Figure 6 and Figure 4. According to the result, it is easy to conclude that a combination of all four image sequences offers the best dice (0.9869 at epoch 300 in the validation phase) and least loss (0.0178 at epoch 300 in validation phase). Our model is reliable and has consistent performance in both the training and validation phase.

In addition, we also apply the models obtained from training with different modality combinations to quantitatively compare the skull stripping performance in the testing phase, and the result shows in Figure 5.The comparison supports the conclusion: the convolutional neural network-based model with the integration of all image modalities offers the best results.

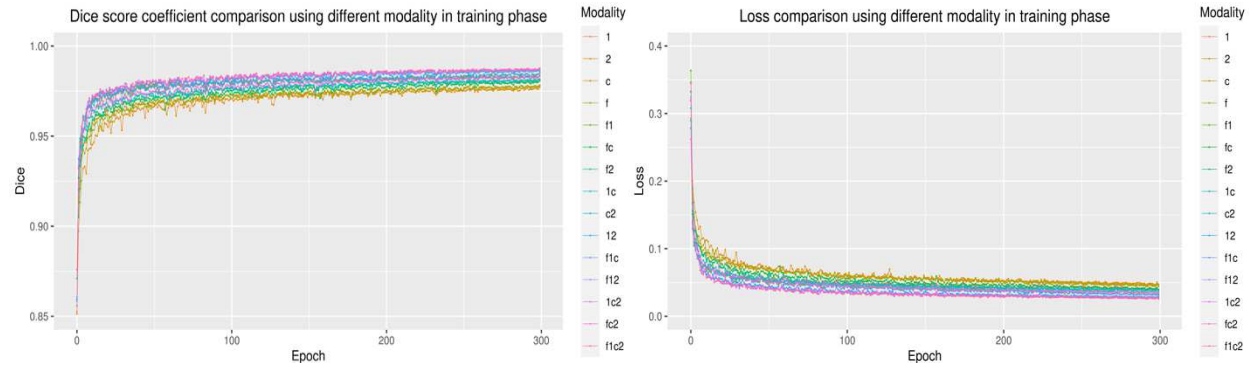

Figure 6. The change of dice (left) and loss (right) in the training phase. In the legend, 1, 2, c, and f represent T1, T2, T1ce, and T2-FLAIR, respectively. For example, f1ce represents that the combination has all four image sequences.

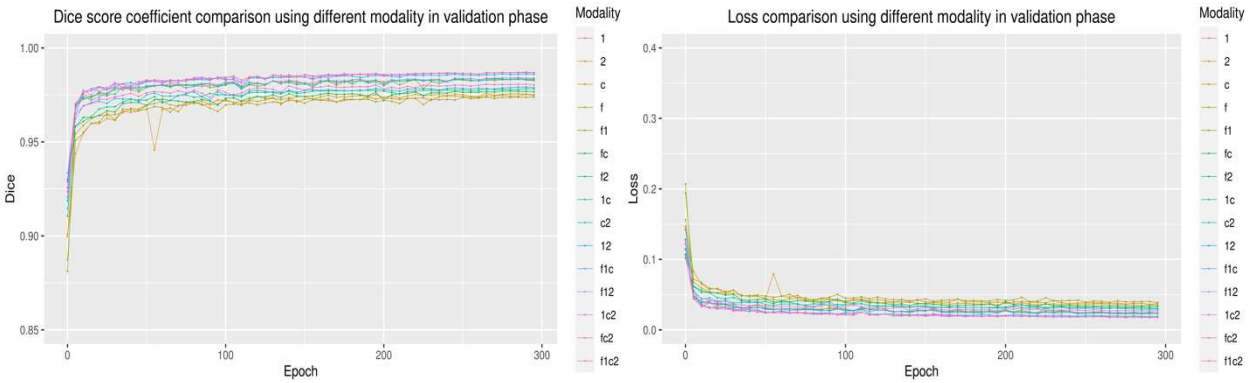

Figure 4. The change of dice (left) and loss (right) in the validation phase. To save training time, we execute the validation part in every 5 epochs. In the legend, 1, 2, c, and f represent T1, T2, T1ce, and T2-FLAIR, respectively. For example, f1ce means that the combination has all four image sequences.

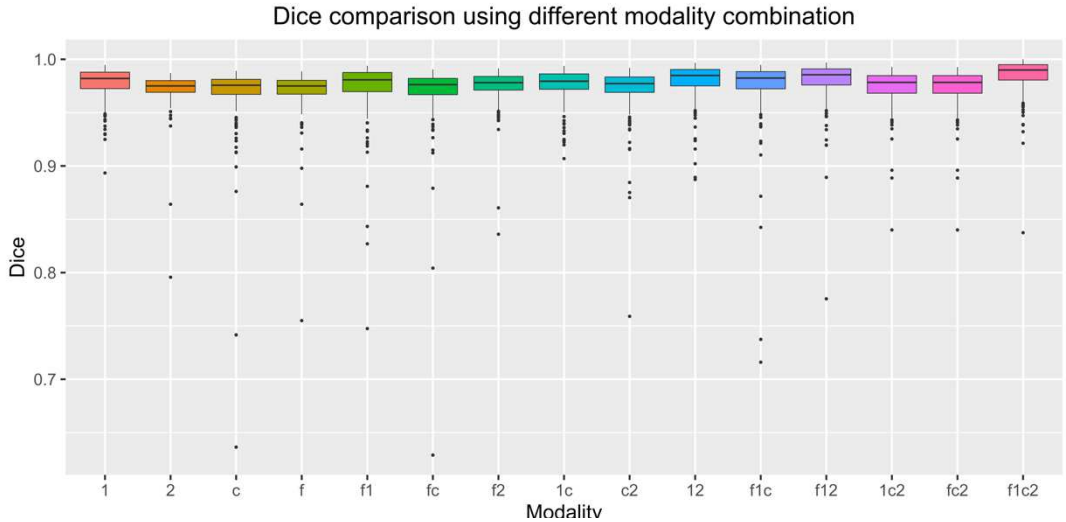

Figure 5. Quantitative dice comparison using different modality combination in the testing phase. In the x axis, 1, 2, c, and f representsT1, T2, T1ce, and T2-FLAIR, respectively. For an example, f1ce means that the combination has all four image sequences.

V. The Proposed Method

Deep neural networks have been becoming successful in many domains and achieve state-of-the-art performance for many applications. Therefore, in the work, we build a deep neural network-based method for skull stripping because of its advantages. The motivation for building a novel skull stripping has three facets. The first one is to process multiparametric brain MRI (mpMRI), which includes T1-weighted (T1), T1-weighted and contrast-enhanced (T1ce), T2-weighted (T2), and T2-fluid-attenuated inversion recovery (T2-FLAIR). The mpMRI offers a better result of skull stripping than that of a single image sequence. Moreover, it is easy to customize for any image sequence combination. Last, the proposed method is generative for all conditional cases, including healthy brain MRI, and pre-/post- operative brain MRI.

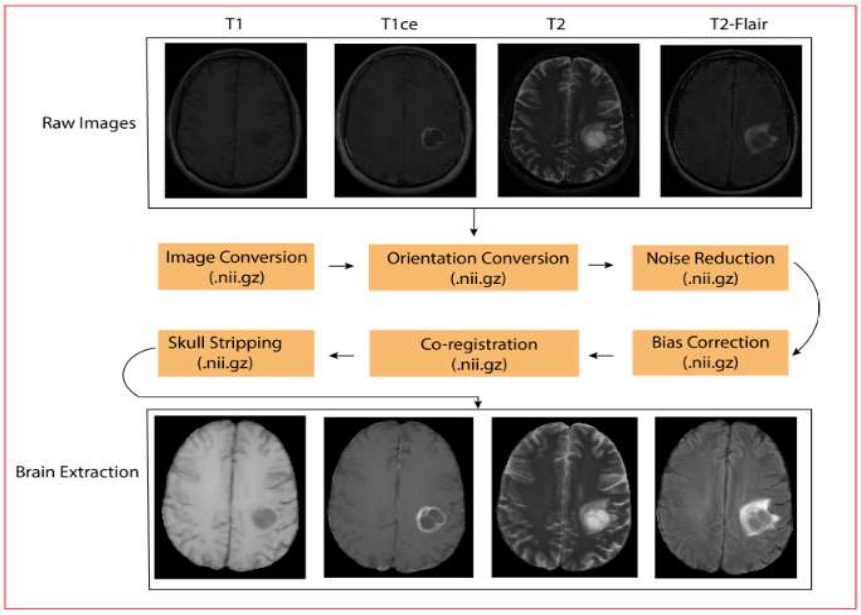

Figure 7. The whole workflow of brain extraction proposed in this work.

The whole workflow of brain extraction is showing *in* Figure 7. Firstly, we convert the raw digital imaging and communication in medicine (.dicom) multiparametric images into a compressed neuroimaging informatics technology initiative (.nii.gz) format, then change the orientation same as to the SRI24 atlas [40]. There are then two optional pre-processing steps: noise reduction and bias correction. Subsequentially, each imaging modality registers to the atlas, so that all image modalities are aligned into the same space. Finally, the co-registered images are fed into the proposed deep neural network model for skull stripping to obtain a binary mask. The co-registered brain extraction is accomplished by multiplying the binary mask to the co-registered images.

The proposed architecture of a deep neural network is illustrated in Figure 8. There are two main parts of the network. The first encoder part is to extract high-dimensional features. The encoder part consists of several convolution blocks and max-pooling blocks. A convolution block is composed of convolution with residual connection, group normalization, and leaky rectified linear unit. Another part is a decoder, which is the opposite function to the encoder. The decoder expands the high-dimensional features to the target segmentation. It consists of convolution blocks and up-sampling blocks. In addition, we design an extra block (convolution block in green). The feature maps add the features from the regular decoder to enforce the training convergency. We name the proposed architecture as an ensemble neural network (EnNet).

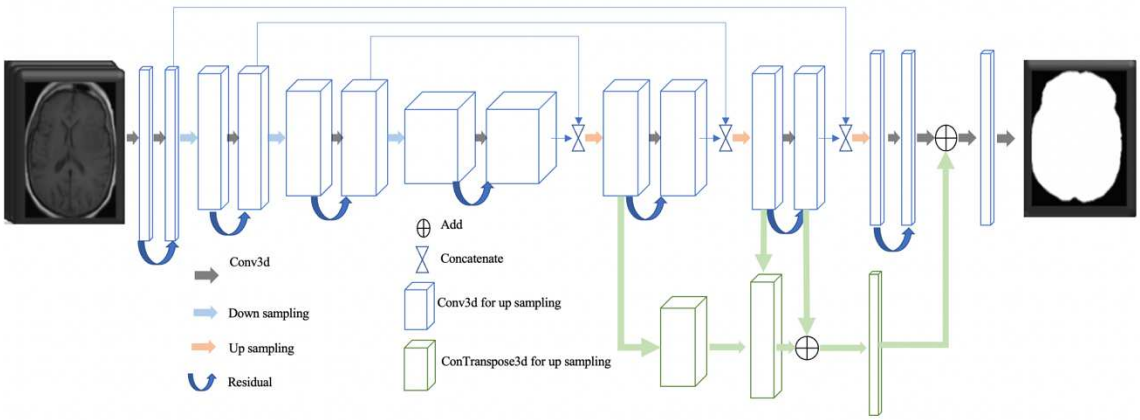

Figure 8. The proposed deep neural network architecture for skull stripping.

VI. Materials and Experiment

All experiments in this study are performed in accordance with relevant guidelines and regulations as approved by the institutional IRB committee at the University of Pittsburgh.

a. Dataset

In this work, we use a total of 815 cases from multi-center for the experiment. Each case has mpMRIs which contain T1-weighted (T1), T1-weighted and contrast-enhanced (T1ce), T2-weighted (T2), and T2-fluid-attenuated inversion recovery (T2-FLAIR). Within the 815 cases, 776 cases are obtained from the University of Pittsburgh Medical Center (UPMC), and the rest of 39 cases are coming from The Cancer Genome Atlas (TCGA), which also collects data from multiple institutes. The size of the image varies among all cases. The image size varies from $256 \times 256 \times 23$ to $512 \times 512 \times 89$, where 23 and 89 is the slice number of each case. For the atlas, the size of the SRI24 is $240 \times 240 \times 155$.

b. Experiment Setup

Before skull stripping, there are several pre-processing steps, including image format conversion, orientation change, noise reduction, bias correction, and co-registration, as details discussed in Section V. In the experiment, all cases are split into training (480 cases), validation (119 cases), and testing dataset (216 cases). In the testing dataset, there are 177 cases and 39 cases from UPMC and TCIA, respectively. More specially, the 177 cases consist of 57, 57, and 63 cases for normal brain, pre-operative, post-operative cases, respectively. The 39 TCIA cases are composed of 20 pre-operative and 19 post-operative MRIs. Note that the training and validation data are obtained from our in-house UPMC, but the testing cases are obtained from both UPMC and TCGA for evaluating the generality of the proposed method.

c. Hyper-parameter Setting

In each iteration, we randomly crop all co-registered MRIs with the size as 160x192x128 because of the limited capacity of the graphics processing unit (GPU). We believe that the cropped image covers the most region-of-interest (ROI). The batch size is set as 1 due to the large patch size and limited GPU memory. The loss function is computed as follows:

$$L = -(ylog(p) + (1-y)log\,(1-p)), \tag{1}$$

where p and y are the class prediction and ground truth (GT), respectively.

We use Adam [41] optimizer with an initial learning rate of $lr_0 = 0.001$ in training phase, and the learning rate (r_i) is gradually decayed by the following:

$$r_i = r_0 * (1 - \frac{i}{N})^{0.9}, \quad (2)$$

where i is epoch counter, and N is a total number of epochs in training.

d. Evaluation Measurements

To quantitatively evaluate the performance of the proposed method, we employ several evaluation metrics in the work, such as dice, precision, recall, false positive rate (FPR), false negative ration rate (FNR), and Hausdorff distance at the 95 percentiles (HD95). They are calculated as follows:

$$Dice = F1 = \frac{TP}{2TP+FN+FP}, \quad (3)$$

$$Precision = \frac{TP}{TP+FP}, \quad (4)$$

$$recall = \frac{TP}{TP+FN}, \quad (5)$$

$$FPR = \frac{FP}{FP+TN}, \quad (6)$$

$$FNR = \frac{FN}{FN+TP}, \quad (7)$$

$$HD95 = percentile(max_{a \epsilon pred} min_{b \epsilon gt}\{d(pred, gt)\}, 95^{th}), \quad (8)$$

Where TP, FN, FP, TN are true positive, false negative, false positive, and true negative, respectively.

Dice is a statistic matrix that measures the similarity of the prediction and ground truth. A value of 1 means that the two groups are identical, and a value of 0 shows no overlap at all between the two groups. The precision indicates how many of the positively classified are relevant. Recall, also known as sensitivity, represents how good a test is at detecting the positives. The Hausdorff distance (HD) measures how far two groups of a metric space are from each other. A smaller value of HD suggests more similarity.

VII. Conclusion

In this work, we propose a 3D convolutional neural network-based method to extract the brain. It is a fully automatic computer-aided method. The proposed method works generally for healthy brain MRIs, and pre-/post- operative brain MRIs with tumors as well. Moreover, the trained model using the proposed method is robust. It is not only applicable for in-house private data, but also for multicenter data. Comparing to the performance of state-of-the-art, the proposed method provides the best result. In addition, we first quantitatively evaluate the impact of skull stripping using different MRI sequences (combination). We conclude that the integration of all multiparametric MRI sequences offers the highest accuracy of brain extraction. In the future, we would like to train the deep learning model with more cases and apply the model to more multicenter data.

DATA AVAILABILITY

The partial datasets generated and/or analyzed during the current study are available in The Cancer Imaging Archive (TCGA) repository (link: https://www.cancerimagingarchive.net). The rest data are privately owned by University of Pittsburgh Medical Center (UPMC).

Author Contributions

LP designed and constructed the experiments and wrote the draft of the manuscript. MA, TN, ZS, KS, KA, and YM verified the ground truth of the experimental dataset and revised the manuscript. CL and EM revised the manuscript. CR supervised the whole project and revised the manuscript.

Reference

[1] Q. T. Ostrom, H. Gittleman, G. Truitt, A. Boscia, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015," *Neuro-oncology,* vol. 20, no. suppl_4, pp. iv1-iv86, 2018.

[2] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images," *Scientific Reports,* vol. 10, no. 1, pp. 1-11, 2020.
[3] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging,* vol. 34, no. 10, pp. 1993-2024, 2014.
[4] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629,* 2018.
[5] S. Pereira, R. Meier, V. Alves, M. Reyes, and C. A. Silva, "Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment," in *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*: Springer, 2018, pp. 106-114.
[6] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis,* vol. 35, pp. 18-31, 2017.
[7] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, 2018, pp. 311-320: Springer.
[8] L. Pei, S. M. Reza, W. Li, C. Davatzikos, and K. M. Iftekharuddin, "Improved brain tumor segmentation by utilizing tumor growth model in longitudinal brain MRI," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017, vol. 10134, p. 101342L: International Society for Optics and Photonics.
[9] S. M. Reza, R. Mays, and K. M. Iftekharuddin, "Multi-fractal detrended texture feature for brain tumor classification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015, vol. 9414, p. 941410: International Society for Optics and Photonics.
[10] P. Kalavathi and V. S. Prasath, "Methods on skull stripping of MRI head scan images—a review," *Journal of digital imaging,* vol. 29, no. 3, pp. 365-379, 2016.
[11] N. F. Kalkers, N. Ameziane, J. C. Bot, A. Minneboo, C. H. Polman, and F. Barkhof, "Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype," *Archives of neurology,* vol. 59, no. 10, pp. 1572-1576, 2002.
[12] R. De Boer *et al.*, "White matter lesion extension to automatic brain tissue segmentation on MRI," *Neuroimage,* vol. 45, no. 4, pp. 1151-1161, 2009.
[13] P. Tanskanen *et al.*, "Hippocampus and amygdala volumes in schizophrenia and other psychoses in the Northern Finland 1966 birth cohort," *Schizophrenia research,* vol. 75, no. 2-3, pp. 283-294, 2005.
[14] H. Rusinek *et al.*, "Alzheimer disease: measuring loss of cerebral gray matter with MR imaging," *Radiology,* vol. 178, no. 1, pp. 109-114, 1991.
[15] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *The Cancer Imaging Archive,* vol. 286, 2017.
[16] J. Kleesiek *et al.*, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage,* vol. 129, pp. 460-469, 2016.
[17] C. Fennema-Notestine *et al.*, "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.
[18] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Transactions on medical imaging,* vol. 12, no. 2, pp. 153-166, 1993.
[19] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," (in eng), *Neuroimage,* vol. 13, no. 5, pp. 856-76, May 2001.
[20] H. K. Hahn and H.-O. Peitgen, "The skull stripping problem in MRI solved by a single 3D watershed transform," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2000, pp. 134-143: Springer.
[21] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE transactions on medical imaging,* vol. 23, no. 4, pp. 447-458, 2004.
[22] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage,* vol. 26, no. 3, pp. 839-851, 2005.
[23] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping,* vol. 17, no. 3, pp. 143-155, 2002.
[24] M. Jenkinson, M. Pechaud, and S. Smith, "BET2: MR-based estimation of brain, skull and scalp surfaces," in *Eleventh annual meeting of the organization for human brain mapping*, 2005, vol. 17, p. 167: Toronto.

[25] J.-X. Liu, Y.-S. Chen, and L.-F. Chen, "Accurate and robust extraction of brain regions using a deformable model based on radial basis functions," *Journal of neuroscience methods,* vol. 183, no. 2, pp. 255-266, 2009.
[26] G. B. Aboutanos, J. Nikanne, N. Watkins, and B. Dawan, "Model creation and deformation for the automatic segmentation of the brain in MR images," *IEEE Transactions on biomedical engineering,* vol. 46, no. 11, pp. 1346-1356, 1999.
[27] K. K. Leung *et al.*, "Brain MAPS: an automated, accurate and robust brain extraction technique using a template library," *Neuroimage,* vol. 55, no. 3, pp. 1091-1108, 2011.
[28] S. F. Eskildsen *et al.*, "BEaST: brain extraction based on nonlocal segmentation technique," *NeuroImage,* vol. 59, no. 3, pp. 2362-2373, 2012.
[29] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* (no. 2). MIT press Cambridge, 2016.
[30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.
[31] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Computational intelligenCe magazine,* vol. 13, no. 3, pp. 55-75, 2018.
[32] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience,* vol. 2018, 2018.
[33] A. Hannun *et al.*, "Deep speech: Scaling up end-to-end speech recognition," *arXiv preprint arXiv:1412.5567,* 2014.
[34] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage,* vol. 170, pp. 446-455, 2018.
[35] L. Pei, L. Vidyaratne, W.-W. Hsu, M. M. Rahman, and K. M. Iftekharuddin, "Brain Tumor Classification Using 3D Convolutional Neural Network," Cham, 2020, pp. 335-342: Springer International Publishing.
[36] Y. Gordienko *et al.*, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*, 2018, pp. 638-647: Springer.
[37] H. Hwang, H. Z. U. Rehman, and S. Lee, "3D U-Net for skull stripping in brain MRI," *Applied Sciences,* vol. 9, no. 3, p. 569, 2019.
[38] S. Thakur *et al.*, "Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training," *NeuroImage,* vol. 220, p. 117081, 2020.
[39] S. P. Thakur *et al.*, "Skull-Stripping of Glioblastoma MRI Scans Using 3D Deep Learning," in *International MICCAI Brainlesion Workshop*, 2019, pp. 57-68: Springer.
[40] T. Rohlfing, N. M. Zahr, E. V. Sullivan, and A. Pfefferbaum, "The SRI24 multichannel atlas of normal adult human brain structure," (in eng), *Hum Brain Mapp,* vol. 31, no. 5, pp. 798-819, May 2010.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.
[42] R. de Boer *et al.*, "Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods," *Neuroimage,* vol. 51, no. 3, pp. 1047-1056, 2010.
[43] J. E. Iglesias, C.-Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE transactions on medical imaging,* vol. 30, no. 9, pp. 1617-1634, 2011.