Skull Stripping for MRI: a deep Convolution Neural Network approach

Kaiyuan Chen
Computer Science Department
University of California, Los Angeles
Los Angeles, United States
chenkaiyuan@ucla.edu

Jingyue Shen
Computer Science Department
University of California, Los Angeles
Los Angeles, United States
brianshen@ucla.edu

***Abstract*—In this work, we developed a deep Convolutional Neural Network(CNN) scheme to perform brain skull stripping on Magnetic Resonance Image(MRI). We analyzed previous works on machine learning including ensembled learning and linear models. By conducting series of experiments, we find weaknesses of popular machine models, like linear models and ensembled learning methods on poor scalability and strong assumption on structure of images. In order to reduce above problems, we propose a CNN approach to reach a higher scalablity that can generalize to and visual accuracy.**

***Index Terms*—Medical Imaging, Machine Learning, Skull Stripping, MRI**

I. Introduction

Computer aided diagnosis based on medical images from MRI(magnetic resonance image) has gained ubiquitous usage for its noninvasive, nondestructive, flexible properties[2]. With the help of FLAIR(Fluid-attenuated inversion recovery), Diffusion-weighted(DW) MRI, people can get an anatomical structure of human soft tissues with high resolution. Especially to satisfy the demand for interior and exterior structure of brain structures, MRI can produce cross-sectional images from different angles, for example, top-down, side-to-side and front-to-back: however, having slices from different angles give a lot of challenges in stripping those tissues which people are interested in, from xtra-cranial or non-brain tissues that has nothing to do with brain diseases such as Alzheimers disease, aneurysm in the brain, arteriovenous malformation-cerebral and Cushings disease and etc[1].

As a preliminary step for further analysis, brain segmentation, i.e. skull stripping, needs both speed and accuracy in practice, which should be considered in any algorithms proposed. By Kalavathi et al. [2], they can be classified into five categories: mathematical morphology-based methods, intensity-based methods, deformable surface-based methods, atlas-based methods, and hybrid methods. However, as we further reviewed on state-of-arts that are vaguely described in hybrid methods, we believe machine learning-based methods should also have its own place in brain segmentation. Machine learning is a broad concept that include many interesting algorithms that we would like to implement and experiment on. For example, Butman introduced a robust machine learning method that detects the brain boundary by random forest[2]. As random forest has high expressive power on voxels of brain boundary, this method can reach an high accuracy robustly. Popular methods like deep learning can also applied. For example, Kleesiek et al.[8] used non-parametric 3D Convolutional Neural Network(CNN) to learn important features and reach the highest Dice score among all the methods we have reviewed. However, as a parametric algorithm, GMM also has its place in brain segmentation. For example, Yunjie et al. developed a skull stripping method with an adaptive gauss mixture model and a 3D mathematical morphology method. The GMM is used to classify brain tissues and to estimate the bias field in the brain tissues [5]. These methods, along with well-implemented libraries such as sklearn[6], Tensorflow[7], are readily available for our use.

Our contribution in this work is as following:

- We conduct a series of experiments on previous works of machine learning based skull stripping, including ensembled learning like random forest and linear models like support vector machine(SVM). Then we analyze their weaknesses from observation.
- To solve problems of previous works, we adopt Convolution Neural Network(CNN) in a scheme similar to autoencoder.
- We manually labelled a range of MRI images. As previous sklearn-based works focus heavily on structure of image(pixel color, position and color of surrounding pixel), we labeled various MRI examples that previous models would fail.

II. Problem Formulation

Problem Definition

We formulate our problem in the following way: given an image as matrix X, we can view it as a sum of skull matrix S and stripped matrix X', with dimension w and h, as such

$$S_{ij} = \begin{cases} X_{ij} & \text{if it is skull} \\ 0 & \text{otherwise} \end{cases}$$

and

$$X'_{ij} = \begin{cases} 0 & \text{if it is skull} \\ X_{ij} & \text{otherwise} \end{cases}$$

and X'_{ij} should be an output of our program. What we want is to optimize the following objective:

$$J(X') = \alpha|X_{ij} - X'_{ij}|_2 + |S_{ij} - X'_{ij}|_2$$

where α is a hyperparameter that can regularize the loss function and prevent overfitting. In this case, because brain stripping is a preprocessing step, we tend to penalize more to keep original brain structure the same, so we need a higher α.

Then there are many ways to solve the problem, both in terms of discriminative models and generative models. For all the baseline models in this paper, one can calculate

$$P(S_{ij} = 0|i, j, X_{ij}, \{X_{mn}\}_{d((m,n),(i,j))<\varepsilon})$$

where d is a self-defined distance metric and ε is a predefined patch size. From a input data perspective, there are three major popular methods

- pixel-based. Choosing $\varepsilon = 0$ and only select features based on itself.
- patch-based. This method is the most popular for its scalability, by which one can choose distance to be Eulcidean distance or Manahttan distance and tune a self-defined ε as a hyperparameter.
- image-based. One simply feeds the whole image in. Larger ε usually requires more computational power, but CNN method proposed in section IV will be image-based.

and latter experiments will show the difference on choosing features and ε.

Data Preparation

TODO

III. Baseline Model

IV. CNN model

Intuition

Autoencoder is an unsupervised learning algorithm that tries to learn a function

$$h(x) \approx x$$

The goal of this algorithm is to learn an efficient encoding of the given set of data so that the original data can be retrieved as flawlessly as possible.

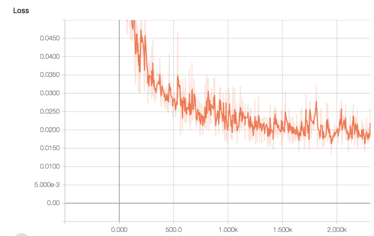
Here we take a similar approach that tries to learn the most representative encoding of the brain structure by compressing and decompressing the given image using Convolutional Neural Network(CNN) model.

Model Setup

We adopt a basic CNN Autoencoder implementation. In encoder part of the model, we choose three convolution layers with three max pooling layers. For the decoder part, we use three convolution layers, three deconvolution layers and an extra reconstruction layer to reconstruct the input image. Note that since we need to reconstruct the skull-stripped images given full images as input, we choose the mean square error(MSE) between stripped images and reconstructed images as loss function, rather than the MSE of input and reconstructed images as used in autoencoders.

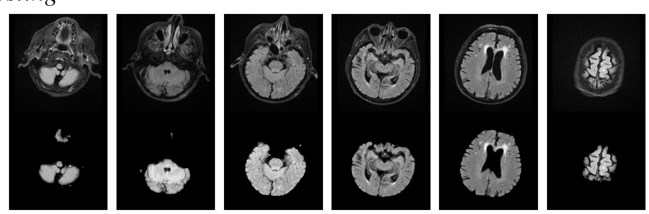
Training

Due to the time consuming manual skull stripping process, we are only able to obtain about 700 original and stripped images as dataset. So we choose batch size to be 15 and the number of epoch to be 50 to train our model. We adopt a roughly 7-2-1 train-validation-test split to choose hyperparameters such as initial learning rate. And we found that learning rate $\alpha = 0.003$ gives best result. The image below is the loss during training process.



After many expriments, we find that our model is able to detect the region of brain in the image, but cannot reconstruct the stripped image with resolution as high as the input. Since skull stripping is a preprocessing step for other diagnosis, which requires high resolution brain images, we use reconstructed images as bitmasks applied to unstripped images.

Testing



The above images are some of our test results. The first row are original unstripped images. And second row are learned stripped images. As shown in those images, our algorithm can perform skull stripping well on various slices of brains. In our dataset, each patient has about 20 brain images. Among all of patients, some of them have 1 to 2 images that are not stripped cleanly. The overall accuracy of our model is around 92%. One thing worth noting is that most of our model's inaccuracy comes from failing to strip a small portion of skull, as shown in the image below. In other words, the core brain parts remain intact, which is desirable for skull stripping.

V. Discussion and Future Work

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.