Brain extraction or whole brain segmentation is an important first step in many of the neuroimage analysis pipelines. The accuracy and robustness of brain extraction, therefore, is crucial for the accuracy of the entire brain analysis process. State-of-the-art brain extraction techniques rely heavily on the accuracy of alignment or registration between brain atlases and query brain anatomy, and/or make strong assumptions about the image geometry; therefore have limited success when these assumptions do not hold or image registration fails. With the aim of designing a learning-based, geometry-independent and registration-free brain extraction tool in this study, we present a technique based on an auto-context convolutional neural network (CNN), in which intrinsic local and global image features are learned through 2D patches of different window sizes. In this architecture three parallel 2D convolutional pathways for three different directions (axial, coronal, and sagittal) implicitly learn 3D image information without the need for computationally expensive 3D convolutions. Posterior probability maps generated by the network are used iteratively as context information along with the original image patches to learn the local shape and connectedness of the brain, to extract it from non-brain tissue. The brain extraction results we have obtained from our algorithm are superior to the recently reported results in the literature on two publicly available benchmark datasets, namely LPBA40 and OASIS, in which we obtained Dice overlap coefficients of 97.42% and 95.40%, respectively. Furthermore, we evaluated the performance of our algorithm in the challenging problem of extracting arbitrarily-oriented fetal brains in reconstructed fetal brain magnetic resonance imaging (MRI) datasets. In this application our algorithm performed much better than the other methods (Dice coefficient: 95.98%), where the other methods performed poorly due to the non-standard orientation and geometry of the fetal brain in MRI. Our CNN-based method can provide accurate, geometry-independent brain extraction in challenging applications. This in-turn may reduce the problems associated with image registration in segmentation tasks.

Index Terms—Brain extraction, Whole Brain Segmentation, MRI, Convolutional Neural Network, CNN, Auto-Context, ACNN.

I. INTRODUCTION

WHOLE brain segmentation, or brain extraction, is one of the first fundamental steps in the analysis of magnetic resonance images (MRI) in advanced neuroimaging applications such as brain tissue segmentation and volumetric analysis [1], longitudinal and group analysis [2], cortical and sub-cortical surface analysis and thickness measurement [3], [4], and surgical planning. Manual brain extraction is time consuming especially in large-scale studies. Automated brain extraction is necessary but its performance and accuracy is critical as the output of this step can directly affect the performance of all next steps. Recently neural networks and deep learning have attracted enormous attention in medical image processing. Brebisson et.al. [5] proposed the SegNet, a convolutional neural network system to segment different parts of the brain. CNN-based methods have also been recently used successfully in tumor segmentation [6], [7], [8], brain lesion segmentation [9], [10], and infant brain image segmentation [11]. In what follows we review the state-of-the-art in whole brain segmentation and the related work that motivated this study. We then introduce a CNN-based method that generates accurate brain extraction.

II. RELATED WORK

Many algorithms have been developed and continuously improved over the past decade for whole brain segmentation, which has been a necessary component of large-scale neuroscience and neuroimage analysis studies. As the usage of these algorithms dramatically grew, the demand for higher accuracy and reliability also increased. Consequently, while fully-automated, accurate brain extraction has already been investigated extensively, it is still an active area of research. Of particular interest is a recent deep learning based algorithm [12] that has shown to outperform most of the popular routinely-used brain extraction tools.

The state-of-the-art brain extraction methods and tools use evolved combinations of image registration, atlases, intensity and edge feature information, and level sets/graph cuts to generate brain masks in MRI images. The majority of these algorithms rely heavily on the alignment of the query images to atlases or make strong assumptions about the geometry, orientation, and image features. Yet the outcome of most of these tools is often inaccurate and involves non-brain structures or cuts parts of the brain. Therefore most of these tools offer options and multiple parameters to set and try, that ultimately make brain extraction a semi-automatic or supervised task rather than fully automatic.
Among brain extraction methods four algorithms that are distributed with the widely-used neuroimage analysis software packages, have been evolved and are routinely used. These are the Brain Extraction Tool (BET) from FSL [13], [14], 3dSkullStrip from the AFNI toolkit [15], the Hybrid Watershed Algorithm (HWA) from FreeSurfer [16], and Robust Learning-Based Brain Extraction (ROBEX) [17]. BET expands a deformable spherical surface mesh model initialized at the center-of-gravity of the image based on local intensity values and surface smoothness. 3dSkullStrip, which is a modified version of BET, uses points outside of the expanding mesh to guide the borders of the mesh. HWA uses edge detection for watershed segmentation along with an atlas-based deformable surface model. ROBEX fits a triangular mesh, constrained by a shape model, to the probabilistic output of a brain boundary classifier based on random forests. Because the shape model alone cannot perfectly accommodate unseen cases, Robex also uses a small free-form deformation which is optimized via graph cuts.

The current methods are prone to significant errors when certain geometric assumptions do not hold, features are not precisely identified, or image registration, which is often not guaranteed to converge to an exact solution, fails. The problems associated with registration-based segmentation, and the recent promising results in neural network based image segmentation motivates further development and use of learning-based, geometry-independent, and registration-free brain image segmentation.

Recently, Kleesiek et. al. [12] proposed a deep learning based algorithm for brain extraction, which will be referred to as PCNN in this paper. PCNN uses seven 3D convolutional layers for voxelwise image segmentation. Cubes of size $53 \times 53 \times 53$ around the grayscale target voxel are used as inputs to the network. In the extensive evaluation and comparison reported in [12], PCNN outperformed state-of-the-art brain extraction algorithms in publicly available benchmark datasets.

In this study we introduce a CNN-based method with three main contributions to further improve brain extraction accuracy. First, instead of using 3D convolutional layers with one window size, we use 2D patches of three different sizes as recently proposed by Moeskops et al. [18]; second, to account for 3D structure, and efficiently learn from 3D information to identify brain voxels from non-brain voxels, we use three parallel pathways of 2D convolutional layers in three different planes (i.e. axial, coronal and sagittal planes); third, we propose an auto-context network architecture that significantly boosts sensitivity in the expense of only a slight decrease in specificity, thus generates accurate brain masks. We discuss the details of our proposed architecture in this paper.

Experimental results in this study show that our method outperformed PCNN and the four widely-used publicly-available brain extraction techniques reviewed above on two benchmark datasets (i.e. LPBA40 and OASIS, described in Section 4.1). On these datasets we achieved significantly higher Dice coefficients by our auto-context CNN-based method, named ACNN, compared to the routinely-used techniques. We also examined the performance of ACNN in the challenging problem of extracting fetal brain from reconstructed fetal brain MRI. In this case we only compared our results to BET and 3dSkullStrip as the other methods were not designed to work with the non-standard orientation and geometry of the fetal brain in MRI. We present the methods in the next section and follow with the experimental results in Section IV and discussion in Section V.

III. Method

A. Network Architecture

The proposed network has 9 types of input features and 9 corresponding pathways which are merged in two levels. Each pathway contains 3 convolutional layers. This architecture segments a 3D image voxel-by-voxel. For all voxels in the 3D image 3 sets of in-plane patches in axial, coronal, and sagittal planes are used. Each set contains three patches with window sizes of $15 \times 15$, $25 \times 25$, and $51 \times 51$. By using these sets of patches with different window size, both local and global features of each voxel are considered during training. Network parameters of learned simultaneously based on orthogonal-plane inputs, so 3D features are learned without using 3D convolution which is computationally expensive.

Figure 1(a) shows the schematic architecture of the parallel 2D pathways for one of the 2D views. In the first layer, twenty four $5 \times 5$ kernels for the patches of size $15 \times 15$ and $25 \times 25$, and $7 \times 7$ kernels for the patches of size $51 \times 51$ are used. After the first convolutional layer, Relu nonlinear function, max pooling with $2 \times 2$ kernel size, and batch normalization is applied. For the second convolutional layer, Relu nonlinear function is used after applying convolutional layer with $32$ convolutional kernels of sizes $3 \times 3$, $3 \times 3$, and $5 \times 5$, for each patch, respectively. In this layer max pooling is only applied to the two largest patches. In the last convolutional layer 48 kernels of size $3 \times 3$ are used. After applying Relu function, max pooling is only performed for the largest patch. The output of the third convolutional layer is connected to a fully connected layer with 256 nodes, with 0.5 dropout for each patch. Then, the nodes for each patch are concatenated and connected to a fully connected layer with 256 nodes combined by 0.5 dropout. Each of these 2D pathways collects the information of a 2D plane.

To combine the information of 2D planes, the outputs of the fully connected layers for each set of in-plane patches are concatenated. This results in 768 nodes in total. These concatenated nodes are subsequently connected to the softmax output layer with 2 nodes (brain and non-brain classes). Figure 1(b) illustrates this step. Based on this architecture, we aim to combine low-level features from patches with context information learned by the network to significantly improve classification accuracy. To this end, we propose an auto-context convolutional neural network by adopting the auto-context algorithm designed in [19].

B. Auto-Context CNN

Assuming $m$ training MRI image pairs $\{(X_j, Y_j), j = 1...m\}$, we add the reshaped 3D input MRI data into the 1D vector $X_j = (x_{j1}, x_{j2}, ..., x_{jn})$ and their corresponding labels into the vector $Y_j = (y_{j1}, y_{j2}, ..., y_{jn})$; where $y_{ji}$ is the...
The optimization, the cross-entropy between the true distribution and the estimated distribution follows the Dirac function, i.e. $\delta(q,p) = \sum_i q(i) \log p(i)$, is minimized. The true distribution follows the Dirac function, i.e. $q(i)$ is 1 for the true label and 0 otherwise. The cost function, therefore, would be:

$$H = - \sum_i \log p(y_i = \text{trueLabel}|X(N_i))$$  \hspace{1cm} (2)

In auto-context CNN, a sequence of classifiers is designed in a way that to train each classifier the posterior probabilities computed by the previous classifier are used as features. More specifically, for each image at step $t$ the pair of $X(N_i), p^{t-1}(N_i)$ is considered as a feature for classification of voxel $i$, where $p^{t-1}(N_i)$ is the posterior probability of voxels around voxel $i$. Algorithm 1 shows how the sequence of weights in the network is computed for the sequence of classifiers. The learned weights are used at test time for classification. The proof of convergence of algorithm 1 is shown in Appendix A.

To illustrate more on the effect of the auto-context algorithm, consider the first convolutional layer of each 2D pathway in the network. Suppose $y$ is an input 3D batch result of concatenating the predicted label patch and data patch, and the information of 2D pathways for 3D segmentation; and c) the auto-context formation of the network to reach the final results. The context information along with multiple local patches are used to learn local shape information from training data and predict labels for test data.

$$x = \sum_{i=1}^{d} W_i * y_i + b$$  \hspace{1cm} (3)

where $W$ is a $k \times k \times d$ weight matrix, $*$ is the 2D convolution operation, $d$ is the depth of the input feature which is 2, and $b$ is the bias. Expanding the summation in equation (3) we have

$$x = W_1 * y_1 + W_2 * y_2 + b$$  \hspace{1cm} (4)

where $W_1$ and $W_2$ are $k \times k$ weight matrices corresponding to the intensity input ($y_1$) and label input ($y_2$), respectively. $W_2$ values are optimized such that they encode information regarding the shape of the brain labels, their respective location, and the connectedness of the labels. During the training of the network at step 0, the weights corresponding to the label input, $W_2$, are assigned much lower values than the weights corresponding to the intensity input (i.e. $W_2 << W_1$)
since the label input carries no information about the image at the beginning. Note that $p_t^1(N_t)$ is constructed with uniform distribution over classes. On the other hand, in the next steps, the weights corresponding to the label input, $W_2$, are assigned higher values than the weights corresponding to the intensity input (i.e. $W_2 > W_1$). Consequently, in testing, the filters corresponding to the predicted labels are more effective than the filters corresponding to intensities.

C. Training

MRI image labels are often unbalanced. For brain extraction the number of non-brain voxels is in average roughly ten times more than the number of brain voxels. The following process was used to balance the training samples: for each training image, 15000 voxels were randomly selected such that 50% of the training voxels were among border voxels. The voxels which had 2 different class labels in a cube of 5 voxels around them were considered border voxels. Of the remaining 50% of samples, 25% were chosen randomly from the brain class and 25% were chosen from the non-brain class.

For training, the cross-entropy loss function was minimized using ADAM optimizer \cite{kingma2014adam}. Three different learning rates were employed during the training. In the first step, the network was trained using a learning rate of 0.001, with 5000 samples for each MRI data pair and 15 epochs. In the second step, learning rate of 0.0001 was used to update the network parameters with another 5000 samples for each MRI data and 15 epochs. Finally, the last 5000 samples for each MRI data and learning rate of 0.00005 were used to update the network parameters.

Figure 1 illustrates the procedure of using algorithm 1. To create patches for each voxel in the network, two sets of features are used; first, patches of different sizes around each voxel are considered as inputs, i.e. $X(N_i)$. Second, exact same patch windows are considered around the posterior probability maps calculated in the previous step, $p_t^{\alpha-1}(N_i)$, as additional sets of inputs. The posterior probabilities are multiplied to the mean of the data intensity to be comparable with data intensities. Concatenating these two 2D features provides 3D inputs to the network in two different domains.

Training was stopped when it reached convergence, i.e. when the change in the cross-entropy cost function became asymptotically smaller than a predefined threshold $\epsilon$:

$$I_t = |H_t - H_{t-1}| < \epsilon$$  (5)

For testing, the auto-context algorithm was used with 2 steps, i.e. $T = 2$. Post-processing of generated brain masks involves Gaussian smoothing with standard deviation of 0.5 followed by applying the connected components algorithm on the brain mask and its inverse \cite{zitovska2003testing}. This warrants a smooth and fully connected brain mask as is the case for most brain extraction tools.

IV. Experiments

A. DataSets

We evaluated our algorithm first on two publicly available benchmark datasets and then on fetal MRI data which exhibits specific challenges such as non-standard, arbitrary geometry and orientation of the fetal brain, and the variability of structures and features that surround the brain. We used two-fold cross-validation in all experiments. The output of all algorithms was evaluated against the ground truth which was available for the benchmark datasets and was manually obtained prior to this study for the fetal MRIs.

The first dataset came from the LONI Probabilistic Brain Atlas Project (LPBA40) \cite{mori2011segmentation}. This dataset consisted of 40 T1-weighted MRI scans of healthy subjects with spatial resolution of $0.86 \times 1.5 \times 0.86$ mm. The second dataset involved the first two disks of the Open Access Series of Imaging Studies (OASIS) \cite{probabilistic04}. This consisted of seventy seven $1 \times 1 \times 1$ mm T1-weighted MRI scans of healthy subjects and subjects with Alzheimer’s disease.

The third dataset contained 75 reconstructed T2-weighted fetal MRI scans. Fetal MRI data was obtained from fetuses scanned at a gestational age between 19 and 39 weeks (mean=30.1, stdev=4.6) on 3-Tesla Siemens Skyra scanners with 18-channel body matrix and spine coils. Repeated multiplanar T2-weighted single shot fast spin echo scans were acquired of the moving fetuses, ellipsoidal brain masks defining approximate brain regions and bounding boxes in the brain region were defined in ITKSNAP \cite{itksnap}, and the scans were then combined through robust super-resolution volume reconstruction by either of the algorithms developed in \cite{itksnap} or \cite{mori2011segmentation} for motion correction and volume reconstruction. Brain masks were manually drawn on the reconstructed images by two experienced segmenters. Manual brain extraction took between 1 to 4 hours per case depending on the age and size of the fetal brain and the quality of the images.

B. Results

To evaluate the performance of the algorithms, Dice overlap coefficient was used to compare the predicted brain mask $P$ with ground truth mask (extracted manually) $R$. The Dice coefficient was calculated as follow:

$$D = \frac{2|P \cap R|}{|P| + |R|} = \frac{2TP}{2TP + FP + FN}$$  (6)

where $TP$, $FP$, and $FN$ are the true positive, false positive, and false negative rates, respectively. We also report the specificity, $\frac{TP}{TP + FN}$, and sensitivity, $\frac{TP}{TP + FP}$, to compare the algorithms.

Figure 2 shows the Dice coefficient, sensitivity, and specificity for the different steps of the training session for fetal MRI in the auto-context CNN algorithm. Significant improvement in the Dice coefficient and sensitivity is observed.

Paired t-test was used to compare the results. The Dice coefficient of the proposed algorithm (ACNN) was significantly higher than BET, 3dSkullStrip, and HWA for LPBA40 and OASIS datasets at $\alpha$ threshold of 0.001 ($p < 0.001$). Compared to Robex, the Dice coefficient of ACNN was significantly higher at $\alpha$ threshold of 0.01 in LPBA dataset but the difference was not significant ($p = 0.5$) in the OASIS dataset. Moreover, it revealed significant differences ($p < 0.001$) between the Dice coefficient of the proposed algorithm with BET and 3dSkullStrip in fetal MRI.
Table I shows the results of the proposed method (ACNN) compared to the other methods on the two benchmark datasets. The results for PCNN were taken from [12]. Our proposed algorithm (ACNN) shows the highest Dice coefficient among all methods, with an increase of about 0.5% over the best performing methods in the LPBA40 dataset. This significant boost in performance was achieved by the architecture of this CNN and the auto-context algorithm which, by incorporating local shape context information along with local patches, allowed a significant increase in sensitivity with only slight decrease in specificity.

The main advantage of our CNN-based method was revealed in the fetal MRI application where the fetal brains were arbitrary oriented and surrounded by a variety of non-brain structures. Figure 3 shows an example, and Table II shows the results of the whole brain segmentation on reconstructed fetal MRI. Only ACNN, BET, and 3dSkullStrip were included in this comparison as the other methods were not designed to work with arbitrary brain orientation in fetal MRI and thus performed poorly. As expected, ACNN performed much better than the other algorithms in this application. The Dice coefficient of the ACNN was approximately 10% higher than the other techniques, and the sensitivity was higher by more than 20%. In fact, as seen in figure 3, the other two algorithms generated conservative brain masks which resulted in high sensitivity (close to 1) but with very low specificity.

The effect of using the auto-context algorithm can be seen in figure 3 where the first network is the network without auto-context. Three different improvements are observed after using auto-context steps. First, the label of the brain voxels considered as non-brain by the first network in the middle of the brain voxels (i.e., false negatives) are changed to brain voxels (yellow arrows). Second, the very small number of the non-brain voxels considered as brain voxels in the first network (i.e., false positives) are changed to non-brain voxels. Third, the second step (auto-context) slightly pushes the edges of the brain to the outside (cyan arrows). These three improvements result in remarkable improvement in sensitivity at the cost of only a slight decrease in specificity. The result is a significant boost in segmentation accuracy also shown by an increase in the Dice overlap coefficient.

Figure 4 shows the box plots of the Dice coefficient, sensitivity, and specificity of the different algorithms on the three datasets. Robex and ACNN worked well on the benchmark datasets. However, Robex could not be used reliably in the fetal dataset because of the geometric assumptions and the use of an atlas. On the other hand, BET and 3dSkullStrip had more relax assumptions thus could be used, albeit with limited accuracy. It should be noted that none of these methods were designed and tested for fetal brain MRI, so it was not expected that they worked well under the conditions of this dataset. Nevertheless, these experiments show the power of our CNN-based method and the advantages of a geometry-independent, registration-free segmentation method.

The influence of the auto-context network is clear in figure 4 when compared to the output of the network in first step (cyan for the result of the first network and blue for the improvement of the results using ACNN). A significant increase in sensitivity at the cost of only slight decrease in specificity was achieved after the auto-context learning steps.

Figure 5 shows logarithmic-scale average absolute error heat map of the different algorithms on the LPBA40 dataset in the MNI atlas space [27]. These maps show where most errors occurred for each algorithm, and indicates that the ACNN performed relatively better than the other methods in this dataset.

V. DISCUSSION

Our proposed Auto-context convolutional neural network architecture outperformed the recent deep learning method [12] and four widely-used brain extraction techniques that were continuously evolved and improved over the past decade due to the significant demand for accurate and reliable automated brain extraction in the neuroscience and neuroimaging communities.

We achieved the highest Dice coefficients and a good sensitivity-specificity trade-off among the techniques examined in this paper. This was achieved by using multiple patch sizes as well as context information in a new CNN architecture.

The main advantage of the CNN-based methods in general, and our method in particular, over the routinely-used state-of-the-art methods is that the CNN-based methods are geometry-independent and registration-free. These techniques are therefore generalizable and can be applied to many applications such as the fetal brain MRI examined in this article. Our technique works on 3D medical images but is efficient due to the use of parallel 2D pathways instead of 3D convolutions. ACNN was easily trained on relatively small number of training datasets as the overlapping image patches already provided large amount of data for training.

With ACNN we overcome one of the persisting challenges in fetal brain MRI processing. The extraction of fetal brain from reconstructed fetal MRI, previously required significant amount of work to correct the masks provided by BET or other level set whole brain segmentation techniques [28]. Atlas-based segmentation methods could not be used as they required alignment to an age-matched atlas, that is also difficult due to the arbitrary orientation of the brain and the inclusion of non-brain material in the absence of brain extraction. Rather than being dependent on an image registration process, the ACNN fetal brain extractions work at the voxel level to mask the fetal brains and prepare them for registration to an atlas space for further analysis.
Despite the challenges raised, our method (ACNN) performed very well and much better than the other methods in this application. The Dice coefficient, surrounded by different tissue or organs such as the amniotic fluid, other fetal body organs such as hands or feet, umbilical cord, or the uterus wall or placenta. Shows the ground truth manual segmentation. As can be seen, fetal brains can be in non-standard arbitrary orientations. Moreover, the fetal head may be surrounded by different tissue or organs such as the amniotic fluid, other fetal body organs such as hands or feet, umbilical cord, or the uterus wall or placenta.

Figure 3. Predicted masks overlaid on the data for fetal brain MRI; the top images show the improvement of the predicted brain mask in different steps of the proposed algorithm. The bottom center and left images show the predicted brain masks using BET and 3dSkullStrip, respectively. The bottom right image shows the ground truth manual segmentation. As can be seen, fetal brains can be in non-standard arbitrary orientations. Moreover, the fetal head may be surrounded by different tissue or organs such as the amniotic fluid, other fetal body organs such as hands or feet, umbilical cord, or the uterus wall or placenta. Despite the challenges raised, our method (ACNN) performed very well and much better than the other methods in this application. The Dice coefficient, sensitivity, and specificity, calculated based on the ground truth for this case, are shown underneath each image in this figure.

In comparison with other methods, in CNN-based methods, the features are learnt during the training step and no hand-crafted features are needed. We used one modality in this study. It is expected that if multiple modalities, such as T1-weighted, T2-weighted, FLAIR or even CT images, are available and used, they result in increased accuracy. The only change in the architecture will be the additional third dimension of the kernel of the first convolutional layer.

### VI. Conclusion

An auto-context convolutional neural network with three parallel 2D pathways was developed for whole-brain segmentation in 3D magnetic resonance images. This novel technique outperformed a recent deep learning method and four widely-used brain extraction methods in two publicly available benchmark datasets and in the very challenging problem of extracting fetal brain from reconstructed fetal MRI. Unlike the current highly evolved brain extraction methods that use a combination of surface models, surface evolutions, and edge and intensity features, CNN-based methods do not use image registration or assume predefined geometric features such as certain orientations. We achieved superior performance without making such assumptions.

### Appendix A

**Theorem 1.** The cross-entropy cost function in algorithm [1] monotonically decreases during the training.

**Proof.** To show that the cross-entropy cost function decreases monotonically we show that the cost at each level will be smaller or at least equal to the cost at previous level. At the arbitrary step $t$.

$$H_t = - \sum_i \log p_i^{(t)}(y_i)$$

$$= - \sum_i \log p_i^{(t)}(y_i|(X_j(N_i), p^{(t-1)}(N_i)))$$

and

$$H_{t-1} = - \sum_i \log p_i^{(t-1)}(y_i)$$

### Table II

| Method  | Dice         | Sensitivity  | Specificity  |
|---------|--------------|--------------|--------------|
| ACNN    | **95.98**±0.008 | **96.83**±0.02 | **99.05**±0.004 |
| BET     | 83.68±0.07   | 73.00±0.1    | 99.91±0.001 |
| 3dSkull | 80.57±0.12   | 69.19±0.16   | **99.97**±0.001 |

**Mean and standard deviation of the scores of different algorithms on the fetal dataset. The results show that highest Dice coefficients were obtained by ACNN compared to BET and 3dSkullStrip among the techniques that could be used in this application.**
Figure 4. Evaluation scores (Dice, sensitivity, and specificity) for three data sets (LPBA40, OASIS, and fetal MRI). Median is displayed in boxplots; blue crosses represent outliers outside 1.5 times the interquartile range of the upper and lower quartiles, respectively. For the fetal dataset the registration-based algorithms were removed due to their poor performance. Those algorithms were not meant to work for images of this kind with non-standard geometry. Overall, these results show that our method (ACNN) made a very good trade-off between sensitivity and specificity and generated the highest Dice coefficients among all methods including the PCNN [12]. The performance of ACNN was consistently superior in the fetal MRI application where the other methods performed poorly due to the non-standard image geometry and features.
Figure 5. Logarithmic-scale absolute error maps of brain extraction obtained from five algorithms on the LPBA40 dataset. This analysis shows that ACNN performed better than the other methods in this dataset.

Also, note that the posterior probability is:

$$p(t)(y_i = k|X(N(i), p^{t-1}(N(i)))) = \frac{e^{f_{y_k}(X(N(i), p^{t-1}(N(i)))}}{\sum_k e^{f_{y_k}(X(N(i), p^{t-1}(N(i))))}$$

(9)

Using $f_{y_k}(X(N_i), p^{t-1}(N_i)) = \log p_i^{(t-1)}(y_i)$ cross-entropy in level $t$ will be equal to cross-entropy in level $t-1$. Since, during the training in step $t$ we are minimizing the cross entropy cost function, $p_i^{(t)}(y_i)$ should at least work better than $p_i^{(t-1)}(y_i)$. Therefore:

$$H_t \leq H_{t-1}$$

(10)

**ACKNOWLEDGMENT**

This study was supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (NIH) grant R01 EB018988. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**REFERENCES**

[1] A. Makropoulos, I. S. Gousias, C. Ledig, P. Aljabar, A. Serag, J. V. Hajnal, A. D. Edwards, S. J. Counsell, and D. Rueckert, “Automatic whole brain MRI segmentation of the developing neonatal brain,” IEEE transactions on medical imaging, vol. 33, no. 9, pp. 1818–1831, 2014.

[2] G. Li, L. Wang, F. Shi, A. E. Lyall, W. Lin, J. H. Gilmore, and D. Shen, “Mapping longitudinal development of local cortical gyriﬁcation in infants from birth to 2 years of age,” The Journal of Neuroscience, vol. 34, no. 12, pp. 4228–4238, 2014.

[3] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, “Automated 3D extraction of inner and outer surfaces of cerebral cortex from MRI,” Neuroimage, vol. 12, no. 3, pp. 346–356, 2000.

[4] C. Clouchoux, D. Kudelski, A. Gholipour, S. K. Warfield, S. Viseur, M. Bouyssi-Kobar, J.-L. Mari, A. C. Evans, A. J. Du Plessis, and C. Limperopoulos, “Quantitative in vivo MRI measurement of cortical development in the fetus,” Brain Structure and Function, vol. 217, no. 1, pp. 127–139, 2012.

[5] A. de Brebisson and G. Montana, “Deep neural networks for anatomical brain segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 20–28.

[6] C. Wachinger, M. Reuter, and T. Klein, “DeepNAT: Deep convolutional neural network for segmenting neuroanatomy,” Neuroimage, 2017.

[7] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in MRI images,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240–1251, 2016.

[8] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” Medical Image Analysis, 2016.

[9] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsi, and R. Tam, “Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1229–1239, 2016.

[10] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efﬁcient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” arXiv preprint arXiv:1603.05959, 2016.

[11] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, “Deep convolutional neural networks for multi-modality isointense infant brain image segmentation,” Neuroimage, vol. 108, pp. 214–224, 2015.

[12] J. Klessie, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, “Deep MRI brain extraction: a 3D convolutional neural network for skull stripping,” Neuroimage, vol. 129, pp. 460–469, 2016.

[13] S. M. Smith, “Fast robust automated brain extraction,” Human brain mapping, vol. 17, no. 3, pp. 143–155, 2002.

[14] 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, vol. 17. Toronto, ON, 2005, p. 167.

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

[16] Y. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam, “A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks,” Cytometry Part A, vol. 56, no. 1, pp. 23–36, 2003.

[17] J. E. Igiehas, 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.

[18] P. Moeskops, M. A. Vereijzer, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Isgum, “Automatic segmentation of MR brain images with a convolutional neural network,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1252–1261, 2016.

[19] Z. Tu and X. Bai, “Auto-context and its application to high-level vision tasks and 3D brain image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 10, pp. 1744–1757, 2010.

[20] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[21] O. Tsang, A. Gholipour, N. Kehtamavaz, K. Gopinath, and R. Briggs, “Comparison of brain masking techniques in functional magnetic resonance imaging,” in Engineering in Medicine and Biology Workshop, 2007 IEEE Dallas, IEEE, 2007, pp. 78–81.

[22] D. W. Shattuck, G. Prasad, M. Mirza, K. L. Narr, and A. W. Toga, “Online resource for validation of brain segmentation methods,” Neuroimage, vol. 45, no. 2, pp. 431–439, 2009.

[23] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. R. Buckner, “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.
[24] P. A. Yushkevich, J. Fiven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[25] A. Gholipour, J. A. Estroff, and S. K. Warfield, “Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI,” *IEEE transactions on medical imaging*, vol. 29, no. 10, pp. 1739–1758, 2010.

[26] B. Kainz, M. Steinberger, W. Wein, M. Kuklisova-Murgasova, C. Malamateniou, K. Keraudren, T. Torsney-Weir, M. Rutherford, P. Aljabar, J. V. Hajnal et al., “Fast volume reconstruction from motion corrupted stacks of 2D slices,” *IEEE transactions on medical imaging*, vol. 34, no. 9, pp. 1901–1913, 2015.

[27] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, B. D. C. Group et al., “Unbiased average age-appropriate atlases for pediatric studies,” *NeuroImage*, vol. 54, no. 1, pp. 313–327, 2011.

[28] A. Gholipour, A. Akhondi-Asl, J. A. Estroff, and S. K. Warfield, “Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly,” *NeuroImage*, vol. 60, no. 3, pp. 1819–1831, 2012.