Segmenting Hippocampus from Infant Brains by Sparse Patch Matching with Deep-Learned Features

Yanrong Guo[1], Guorong Wu[1], Leah A. Commander[2], Stephanie Szary[3],
Valerie Jewells[2], Weili Lin[1], and Dinggang Shen[1,*]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA
[2] School of Medicine, University of North Carolina at Chapel Hill, NC, USA
[3] Duke University Hospital, NC, USA
dgshen@med.unc.edu

Abstract. Accurate segmentation of the hippocampus from infant MR brain images is a critical step for investigating early brain development. Unfortunately, the previous tools developed for adult hippocampus segmentation are not suitable for infant brain images acquired from the first year of life, which often have poor tissue contrast and variable structural patterns of early hippocampal development. From our point of view, the main problem is lack of discriminative and robust feature representations for distinguishing the hippocampus from the surrounding brain structures. Thus, instead of directly using the predefined features as popularly used in the conventional methods, we propose to learn the latent feature representations of infant MR brain images by *unsupervised deep learning*. Since deep learning paradigms can learn low-level features and then successfully build up more comprehensive *high-level* features in a layer-by-layer manner, such hierarchical feature representations can be more competitive for distinguishing the hippocampus from entire brain images. To this end, we apply Stacked Auto Encoder (SAE) to learn the deep feature representations from both T1- and T2-weighed MR images combining their complementary information, which is important for characterizing different development stages of infant brains after birth. Then, we present a sparse patch matching method for transferring hippocampus labels from multiple atlases to the new infant brain image, by using deep-learned feature representations to measure the inter-patch similarity. Experimental results on 2-week-old to 9-month-old infant brain images show the effectiveness of the proposed method, especially compared to the state-of-the-art counterpart methods.

1 Introduction

During the first year of life, human brains undergo rapid tissue growth and postnatal development. The ability to accurately characterize structural changes from MR images during this period is indispensable for shedding new light upon the exploration of brain development and also the early detection of neurodevelopmental disorders. In many imaging-based early brain development studies, hippocampus is of particular

[*] Corresponding author.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 308–315, 2014.

interest since hippocampus plays an essential role in learning and memory functions. Therefore, accurate and efficient hippocampus segmentation methods for infant brain images is highly demanded in many imaging-based neuroscience studies [1].

Although many hippocampus segmentation methods have achieved success in adult brains and the pediatric brains after 2-year-old [2], segmenting the hippocampus from infant brain images acquired in the first year of life is still challenging. The difficulties include: (1) poor tissue contrast, (2) complex shape/appearance patterns of the hippocampus on MR images, and (3) variable adjacent large structures around the small hippocampus. Therefore, the conventional handcrafted features, such as Haar and HoG features, fail to segment the hippocampus from infant brain images. Although many learning-based approaches can be borrowed to improve the feature discrimination power for hippocampus segmentation, most of them are supervised learning approaches, which often have limited performance when a large set of annotated data (i.e., manually labeled hippocampi) are not available.

In this paper, we address the above challenges by using unsupervised deep learning [3] to directly infer the intrinsic feature representations from the training infant images. In this way, we avoid the dilemma of relying on the manual annotated data. Another advantage of deep learning is that it can infer the hierarchical feature representation in a layer-by-layer manner, i.e., first inferring the *low-level* features and then building up comprehensive *high-level* features based on the learned low-level features. Thus, such hierarchal (local to global) feature representation can be more competitive in characterizing each point of the infant brain image than other widely-used unsupervised learning methods (e.g., PCA and sparse dictionary learning) which can learn only a single-layer feature representation.

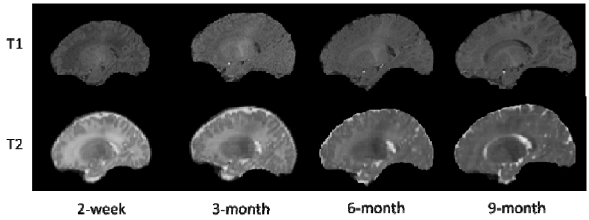


Fig. 1. Typical brain images acquired from 2-week-old to 9-month-old infant. T1- and T2-weighted MR images are provided in the top and bottom rows, respectively.

Since the characteristics of white matter (WM), gray matter (GM), and cerebral-spinal fluid (CSF) change dynamically in the first year of life (as shown in Fig. 1), we use a deep learning technique to independently learn the intrinsic feature representations for three distinct phases, i.e., infantile (birth~5 months old), isointense (6-10 months old), and adult-like (after 10 months old) [4]. Specifically, for each brain development stage, we first collect a large number of image patches from the training infant images at the respective age. Then, we apply Stacked Auto Encoder (SAE) to the collected 3D image patches for inferring the intrinsic hierarchical feature representations. Since T1- and T2-weighted MR images are commonly acquired for each subject and can also provide complementary information, we use both modalities for learning features. Finally, we present a sparse patch matching method for transferring hippocampal labels from multiple atlases [5] to each new infant brain image, by using deep-learned feature representations for measuring inter-patch similarity. In experiments, we have comprehensively evaluated the performance of our method on 10 infant brain subjects acquired from 2-week-old to 9-month-old, obtaining much better results than the state-of-the-art counterpart methods.

2 Method

2.1 Learning Hierarchical Feature Representation by SAE

Our goal here is to use a deep learning technique to infer the intrinsic feature representation for any 3D image patch in the infant MR images. We assume T1- and T2-weighed MR images of same subject are already aligned.[1] Thus, from all training images of different subjects, we can collect a set of paired patches, such as N pairs of $l \times l \times l$ image patches[2], with each pair including a $l \times l \times l$ patch from T1-weighted MRI and another $l \times l \times l$ patch from T2-weighted MRI. Then, we arrange each pair into a column vector, i.e., $\vec{x}_n \in R^L$, $n = 1, \dots, N$, where $L = l \times l \times l \times 2$. In the following, we will describe how to use SAE to learn the intrinsic hierarchical feature representation for each $\vec{x}_n$, by first introducing a single-layer auto encoder (AE).

Single-layer AE: AE consists of two components: the encoder and the decoder. The encoder step seeks for a nonlinear mapping to project the high-dimensional observed data (input units) $\vec{x}_n$ into a low-dimensional code (feature representation). The decoder step aims to recover the observed data (input units) from the low-dimensional code with minimal reconstruction error. Specifically, **in the encoder step**, given the observed data $\vec{x}_n$, the AE maps it to an M-dimensional activation vector, $\vec{h}_n = [h_n(m)]_{m=1}^M$, $\vec{h}_n \in R^M$, $M < L$, through a deterministic mapping, i.e., $\vec{h}_n = \sigma(\boldsymbol{W}\vec{x}_n + \vec{b}_1)$, where the weight/mapping matrix $\boldsymbol{W} \in R^{M \times L}$ and the bias vector $\vec{b}_1 \in R^M$ are the encoder parameters. Here, σ is the logistic sigmoid function, i.e., $\sigma(a) = (1 + \exp(-a))^{-1}$. It is worth noting that the activation vector $\vec{h}_n$ in the hidden layer (with M nodes) is considered as the *low-dimension* feature representation for the input *high-dimension* observed data $\vec{x}_n$. **In the decoder step**, the activation vector $\vec{h}_n$ is then decoded to a vector $\vec{\hat{x}}_n \in R^L$, which approximately reconstructs the input observed data $\vec{x}_n$ by another deterministic mapping, i.e., $\vec{\hat{x}}_n = \sigma(\boldsymbol{W}^T\vec{h}_n + \vec{b}_2) \approx \vec{x}_n$, where $\boldsymbol{W}^T \in R^{L \times M}$ is the transpose of matrix $\boldsymbol{W}$ and $\vec{b}_2 \in R^L$ is the bias vector.

Sparse constraint upon the M hidden nodes in AE can often lead to a small set of more interpretable features. By regarding the m-th hidden node as being “active” if

[1] Since T1- and T2-weighted images were acquired from the same subject at the same time, it is easy to affine align them.

[2] We use a random sampling strategy to collect image patches in a bounding box that covers all possible locations of hippocampus.

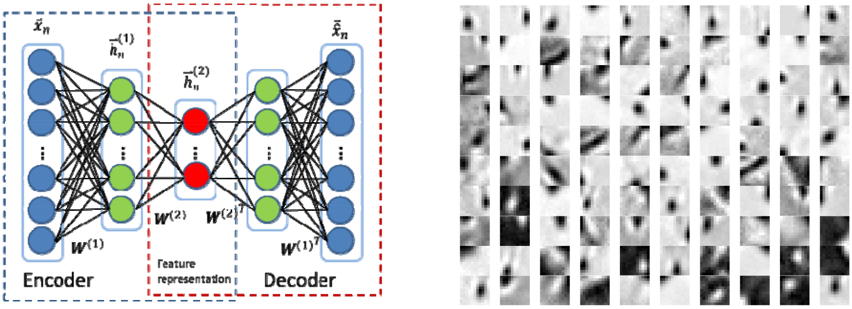
$h_n(m)$ is close to 1, or "inactive" if $h_n(m)$ is close to 0, we can use the sparsity constraint to require most of the hidden nodes to remain "inactive" for each training input data $\vec{x}_n$. Specifically, the Kullback-Leibler (KL) divergence can be used to impose the sparsity constraint to each hidden node (i.e., the m-th hidden node) by enforcing the average activation over the whole training data, defined as $\bar{\rho}_m = \sum_{n=1}^{N} h_n(m)$, to be close to a small value ρ (which we set $\rho = 0.15$ in the experiments below):

$$KL(\rho|\bar{\rho}_m) = \rho \log \frac{\rho}{\bar{\rho}_m} + (1-\rho) \log \frac{1-\rho}{1-\bar{\rho}_m} \tag{1}$$

By integrating the sparsity constraint in the minimization of reconstruction errors between $\vec{x}_n$ and $\hat{\vec{x}}_n$, the objective function of a single-layer AE can be formulated as:

$$\{\boldsymbol{W}, \vec{b}_1, \vec{b}_2\} = \arg\min_{\{\boldsymbol{W}, \vec{b}_1, \vec{b}_2\}} \frac{1}{N} \sum_{n=1}^{N} \left\| \hat{\vec{x}}_n - \vec{x}_n \right\|_2^2 + \beta \sum_{m=1}^{M} KL(\rho|\bar{\rho}_m) \tag{2}$$

where β controls the strength of sparsity constraint.

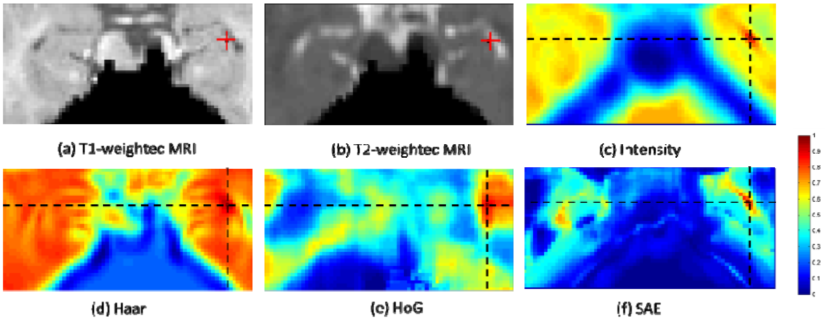


(a) (b)

Fig. 2. (a) A SAE with two hidden layers. (b) Typical mapping matrix $\boldsymbol{W}$ learned by SAE.

SAE for Hierarchical Feature Representation: A single AE is limited in what it can present, since the model is shallow in learning. The power of deep learning emerges when several AEs are stacked to form a Stacked Auto Encoder (SAE), where each AE becomes a building block in the deep learning model. Specifically, in SAE, the input of a high-level AE is the output from the low-level AE in the previous layer. Fig. 2(a) shows a typical SAE with two layers. In this figure, the first-layer AE treats input vector $\vec{x}_n$ (blue circles) as the input and produces the activation vector $\vec{h}_n^{(1)}$ (green circles). Then, the second-layer AE uses the activation vector $\vec{h}_n^{(1)}$ (the output of first-layer AE) as the input to produce the new activation vector $\vec{h}_n^{(2)}$ (red circles), which can be used as the low-dimensional feature representation. To train such multi-layer network, we follow the layer-wise greedy training procedure by **1)** first training the AE in each layer, where the high-level AE uses the output of the low-level AE as the input; **2)** constructing SAE by stacking the AE in each layer, with the higher-layer AE nested within the lower-layer AE; and **3)** training the entire deep network by a gradient based optimization method to further refine the parameters in SAE [3].

Fig. 2(b) shows some typical weight/mapping matrix $\boldsymbol{W}$ learned in the first-layer of SAE, where we only show one slice from each 3D mapping. Furthermore, in Fig. 3, we demonstrate the enhanced discriminative power of the deep-learned feature representations by SAE over the other handcrafted features (Haar and HoG). Specifically, Figs. 3(a) and 3(b) show the T1- and T2-weighted intensity images, respectively, where the reference point (at the boundary of hippocampus) is denoted by red cross. Figs. 3(c-f) compare the discrimination performances by using simple image intensity (Fig. 3(c)), Haar features (Fig. 3(d)), HoG features (Fig. 3(e)), and deep-learned feature representations by SAE (Fig. 3(f)), all computed from both T1- and T2-weighted MRI using $11 \times 11 \times 11$ patch. Here, each similarity map is computed by comparing the respective features of the reference point (indicated by red cross) w.r.t. all other points in the image. It is obvious that the deep-learned feature representations offer the best discrimination performance.


(a) T1-weightec MRI

(b) T2-weightec MRI

(c) Intensity

(d) Haar

(c) HoG

(f) SAE

Fig. 3. Maps of similarities between the reference point (red cross) and all other points in the image, obtained by 4 different feature representations such as (c) simple image intensity, (d) Haar, (e) HoG, and (f) SAE. All the similarity maps are normalized into [0 1].

2.2 Sparse Patch Matching for Robust Infant Hippocampus Segmentation

Recently, multi-atlas based segmentation is widely used to deal with the high structural variations in the population. In our application, we first align all P atlas images $\{I_p, p = 1, \dots, P\}$ as well as their hippocampus label maps $\{G_p, p = 1, \dots, P\}$ onto the target image I_s. Note that, for clarity, each I_p (or I_s) denotes a pair of self-aligned T1- and T2-weighted MR images for each subject. It is worth noting that we apply only affine registration between each atlas image and the target image, since it is very difficult to non-rigidly register infant images in the first year of life due to dynamic anatomical and appearance changes. To alleviate the possible misalignment, we adopt a patch-based label fusion technique to determine the label (hippocampus or non-hippocampus) for each target image point, by calculating the patch-wise similarity between target image patch and each atlas image patch. Specifically, to determine the label of a particular point v in the target image I_s, we first extract an image patch centered at v. Then, we search for a set of atlas image patches within certain searching neighbor $\mathbb{N}(v)$ across all registered atlas images.

Next, we extract the intensity values from both T1- and T2-weighted MRI within each image patch, and further obtain the deep-learned feature representation (i.e., the low-dimension activation vector in the middle layer of SAE) through the encoder component of SAE. Here, we use $\vec{f}_s(v)$ to denote the deep-learned feature representation for the target image patch, and further arrange all deep-learned feature representations of all atlas image patches into a matrix $\boldsymbol{A}$ (column by column). To achieve robust hippocampus segmentation, we further enforce the sparsity constraint upon the weighting vector $\vec{\alpha}_v = [\alpha_v(p,u)]_{u=1,\ldots,|\mathbb{N}(v)|;\, p=1,\ldots,P}$, where each element in $\vec{\alpha}_v$ denotes the contribution of label carried by a particular atlas image patch in label fusion. Thus, our goal finally turns to finding the optimal weighting vector $\vec{\alpha}_v$ that can minimize the difference between the target feature representation $\vec{f}_s(v)$ and the linearly combined feature representation $\boldsymbol{A}\vec{\alpha}_v$ from all atlas image patches. The overall energy function can be defined as below:

$$\vec{\alpha}_v = \arg\min_{\vec{\alpha}_v} \frac{1}{2} \left\| \vec{f}_s(v) - \boldsymbol{A}\vec{\alpha}_v \right\|_2^2 + \mu \|\vec{\alpha}_v\|_1 \quad \text{s.t. } \vec{\alpha}_v > 0 \tag{3}$$

where μ controls the strength of sparsity constraint on the weighing vector $\vec{\alpha}_v$. We use an optimization method in [6] to solve the above sparse representation problem. After obtaining the optimal weighing vector $\vec{\alpha}_v$, the final likelihood $Q_s(v)$ on the target image point v of I_s can be determined by:

$$Q_s(v) = \frac{\sum_{p=1}^{P} \sum_{u \in \mathbb{N}(v)} \vec{\alpha}_v(p,u) \times G_p(u)}{\sum_{p=1}^{P} \sum_{u \in \mathbb{N}(v))} \vec{\alpha}_v(p,u)} \tag{4}$$

Given the likelihood map for hippocampus, we further apply the level sets method to outline the hippocampus boundary and derive the final segmentation.

3 Experimental Results

In the experiments, MR images of 10 infant subjects acquired from a Siemens head-only 3T scanner are used. In each subject, both T1- and T2-weighted MR images were acquired in four data sets at 2 weeks, 3 months, 6 months and 9 months of age. T1-weighted MR images were acquired with 144 sagittal slices at a resolution of $1 \times 1 \times 1mm^3$, while T2-weighted MR images were acquired with 64 axis slices at resolution of $1.25 \times 1.25 \times 1.95mm^3$. For each subject, the T2-weighted MR image is aligned to the T1-weighted MR image at the same age and then further resampled to $1 \times 1 \times 1mm^3$. In the pre-processing step, skull stripping and bias-field correction is applied to each image. The manual segmentations of the hippocampal regions for all 10 subjects are used as ground-truth for evaluation.

We set parameters for the unsupervised deep feature learning as below. The patch size is set to $11 \times 11 \times 11$ considering the balance between computation time and discriminative power, and 4 layers are employed in the SAE. The number of units in each layer of SAE is 800, 400, 200 and 100, respectively. Thus, the final dimensionality of deep-learned feature representation is 100. The target activation ρ for the

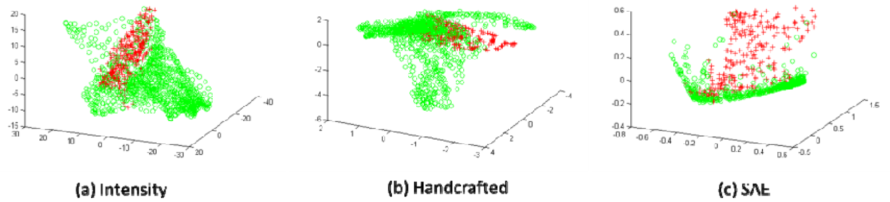


Fig. 4. Scatter plots of samples from the same subject using 3 different feature representations. Red cross and green circle denote hippocampus and non-hippocampus samples respectively

Table 1. Mean and standard deviation of Dice ratio (in %) for the segmenations obtained by three different feature representation methods

Method	2-Week	3-Month	6-Month	9-Month	Overall
Intensity	59.0±12.9	68.5±12.6	69.4±9.4	67.7±12.6	66.1±12.2
Handcrafted	57.1±9.6	61.9±11.5	63.2±8.0	66.1±5.0	62.1±9.1
SAE	**62.3±7.5**	**71.9±1.5**	**71.8±3.9**	**74.6±2.9**	**70.2±6.4**

hidden units is set to 0.15 as mentioned before, and the sparsity penalty β is set to 0.1. Finally, the Deep Learning Toolbox [7] is used for training our SAE framework. Before deploying the sparse patch matching for label propagation, a linear registration (using FLIRT) [8] is used to align all atlas images to the target image. The experiments are conducted in a *leave-one-subject-out* manner.

Our method is compared with two other methods using either simple intensity features or handcrafted features, but using the *same* sparse patch-based labeling step for all comparison methods. Specifically, for the case of using the handcrafted features, we include three popular features, i.e., Haar, HoG, and gradients. Fig. 4 shows the distributions of feature vectors from image patches in the same subject, by each of the three methods. Note that, for the purpose of visualization, the dimensionality of each type of feature vectors is reduced to three by using PCA. This can be seen in Fig. 4 where the deep-learned feature representation can well-separate hippocampus voxels from other voxels, while the other two methods are unable to do so.

For quantitatively evaluating the hippocampus segmentation results, both mean and standard deviation of Dice ratios for all three methods are listed in Table 1 from 2-week-old to 9-month-old data sets, along with the overall performance. After applying paired t-test, our method achieves significant improvement ($p<0.013$) over all other methods in terms of overall Dice ratio. Fig. 5 further shows some typical probability maps and the final segmentations by the three methods.

4 Conclusion

In this paper, we have presented a novel method for segmenting the hippocampus from the infant brain images acquired from the first year of life. Specifically, we address this challenging problem by using the unsupervised deep learning technique to infer the intrinsic hierarchical feature representations for infant hippocampi. By integrating the deep-learned feature representations with the state-of-the-art sparse patch-based label fusion paradigm, we developed a novel hippocampus segmentation method and achieved much better performance than the counterpart methods. In the future, we will improve our segmentation method by developing a unified framework for jointly segmenting hippocampi in all ages, thus achieving both accuracy and longitudinal consistence for all segmented hippocampus results.

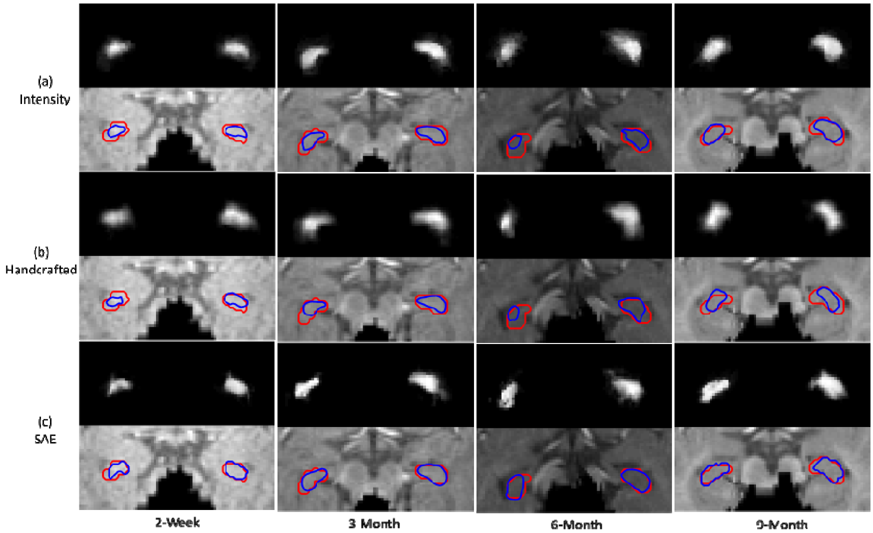


Fig. 5. Typical hippocampus probability maps and segmentation results for 2-week-old to 9-month-old brain images, produced by three different methods. Red contours indicate the manual ground-truth segmentations, and blue contours indicate the automatic segmentations.

References

1. Gousias, I.S., Edwards, A.D., Rutherford, M.A., et al.: Magnetic Resonance Imaging of the Newborn Brain: Manual Segmentation of Labelled Atlases in Term-Born and Preterm Infants. NeuroImage 62(3), 1499–1509 (2012)
2. Jorge Cardoso, M., Leung, K., Modat, M., et al.: Steps: Similarity and Truth Estimation for Propagated Segmentations and Its Application to Hippocampal Segmentation and Brain Parcelation. Medical Image Analysis 17(6), 671–684 (2013)
3. Hinton, G.E., Salakhutdinov, R.R.: Reducing the Dimensionality of Data with Neural Networks. Science 313(5786), 504–507 (2006)
4. Dietrich, R., Bradley, W., Zaragoza, E.T., et al.: MR Evaluation of Early Myelination Patterns in Normal and Developmentally Delayed Infants. American Journal of Roentgenology 150(4), 889–896 (1988)
5. Liao, S., Gao, Y., Lian, J., et al.: Sparse Patch-Based Label Propagation for Accurate Prostate Localization in CT Images. IEEE Transactions on Medical Imaging 32(2), 419–434 (2013)
6. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009)
7. `https://Github.Com/Rasmusbergpalm/Deeplearntoolbox`
8. Jenkinson, M., Bannister, P., Brady, M., et al.: Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage 17(2), 825–841 (2002)