

Unsupervised Deep Learning for Hippocampus Segmentation in 7.0 Tesla MR Images

Minjeong Kim, Guorong Wu, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill

Abstract. Recent emergence of 7.0T MR scanner sheds new light on the study of hippocampus by providing much higher image contrast and resolution. However, the new characteristics shown in 7.0T images, such as richer structural information and more severe intensity inhomogeneity, raise serious issues for the extraction of distinctive and robust features for accurately segmenting hippocampus in 7.0T images. On the other hand, the hand-crafted image features (such as Haar and SIFT), which were designed for 1.5T and 3.0T images, generally fail to be effective, because of the considerable image artifacts in 7.0T images. In this paper, we introduce the concept of unsupervised deep learning to learn the hierarchical feature representation directly from the pre-observed image patches in 7.0T images. Specifically, a two-layer stacked convolutional Independent Subspace Analysis (ISA) network is built to learn not only the intrinsic low-level features from image patches in the lower layer, but also the high-level features in the higher layer to describe the global image appearance based on the outputs from the lower layer. We have successfully integrated this deep learning scheme into a state-of-the-art multi-atlases based segmentation framework by replacing the previous hand-crafted image features by the hierarchical feature representations inferred from the two-layer ISA network. Promising hippocampus segmentation results were obtained on 20 7.0T images, demonstrating the enhanced discriminative power achieved by our deep learning method.

1 Introduction

Numerous methods have been developed for hippocampus segmentation in MR images [1-6], since accurate labeling of hippocampus is significant for study of many neurological diseases, including Alzheimer's disease. However, due to the tiny size of hippocampus ($\approx 35 \times 15 \times 7 \text{mm}^3$) and also the complexity of surrounding structures, the accuracy of hippocampus segmentation is limited by the poor imaging contrast and resolution (i.e., often with the voxel size of $1 \times 1 \times 1 \text{mm}^3$).

Recently, the development of high-resolution imaging technique makes a rapid progress in hippocampus segmentation. For example, in the 7.0T scanner, much more detailed hippocampal structures can be observed, compared to the 3.0T scanner [7]. However, it is not straightforward to apply the existing segmentation methods (developed for 1.5T or 3.0T) to 7.0T images, since the image content is significantly different, in terms of rich structural information and severe intensity inhomogeneity in 7.0T images. The typical examples of 7.0T and 1.5T MR images are shown in Fig. 1, where we can see the obvious difference.

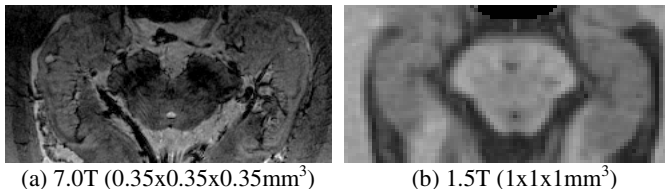


Fig. 1. Large difference between 7.0T (a) and 1.5T (b) MR images

Thus, it is not difficult to notice that the conventional segmentation methods will encounter difficulties in segmenting hippocampus from 7.0T MR images, because of (1) severe intensity inhomogeneity in the 7.0T that can adversely affect the feature consistency of similar anatomical structures; (2) high signal-to-noise ratio (SNR) which brings forth plenty of anatomical details at the expense of troublesome image noise; (3) incomplete brain volume (i.e., with only a segment of brain, considering the practical issue during image acquisition).

Many learning-based methods [8-12] can be used for automatically segmenting hippocampus in 1.5T or 3.0T images. For example, Adaboost is able to learn image features by building a sequence of weak classifiers. It is assumed in this approach that the features (e.g., Haar or SIFT) are general enough to represent the input images. However, it is not straightforward to apply this approach to the 7.0T images which always have various troublesome artifacts (as shown in Fig. 1(a)). Furthermore, it is not guaranteed to obtain robust classifiers by the existing learning-based approaches, unless a large number of manual segmented samples are available for supervised learning.

Inspired by the recent success of deep learning [13, 14] in the field of computer vision and machine learning, we introduce the concept of deep learning to perform unsupervised learning directly on the 7.0T images with following reasons: (1) Deep learning is a unsupervised learning method, which leverages the plethora of unlabeled data for training; (2) Deep learning is able to provide the hierarchical feature representation for each image patch and ensure the discriminative power and robustness for the learned image features; (3) The success of deep learning in analyzing natural images also motivates us to apply it to 7.0T images with its rich image content as the natural images.

Specifically, we apply a two-layer stacked convolutional ISA [15] to learn the hierarchical feature representations from the image patches extracted from 7.0T MR images. To segment hippocampus for a new subject, we first extract image patch for each point and calculate the response of the extracted patch through the trained ISA network. The response vector is regarded as the intrinsic morphological representation for characterizing the underlying image point/patch. Then, we incorporate the learned hierarchical feature representations into a state-of-the-art multi-atlas segmentation framework to label hippocampus in the 7.0T images. Our proposed hippocampus segmentation method with unsupervised deep learning achieves more accurate labeling results than the counterpart with hand-crafted features, as it provides a better way to represent image features in the 7.0T MR images.

2 Method

The goal of learning-based hippocampus segmentation methods is to accurately label each image point $x \in \Omega$ in a new subject into either positive (i.e., hippocampus) or negative (i.e., non-hippocampus). Generally, a set of image features are extracted from a neighborhood of x , which is used as the morphological pattern to label point x .

Although 7.0T image displays plenty of image details around hippocampus, it also introduces severe noise and intensity inhomogeneity compared to the lower-resolution 1.5T or 3.0T images, which raises critical issues of using conventional hand-crafted image features for labeling. By taking Haar features as the example, we examine its discriminative power in representing image patches at different locations of 7.0T image. As shown in Fig. 2, image patches b and c belong to the hippocampus, but image patch a does not. The zoom-in views of these three patches are displayed in Fig. 2(a)-(c). Although the image contents are quite different among these patches, the responses from Haar filters are very similar, as shown in Fig. 2(d)-(f), where each column represents the responses from a set of Haar filters and each row denotes one point in the image patch. According to this observation, we can predict that the hand-crafted Haar features will be not distinctive enough to guide hippocampus segmentation in the 7.0T MR images.

Inspired by the recent success of deep learning [13, 14] in recognizing natural images, we introduce the concept of deep learning to directly extract the hierarchical feature representation of image patches in the 7.0T images (which contains the complex patterns as the natural images). Then, we incorporate the hierarchical feature representation into a state-of-the-art learning-based segmentation framework, for improving hippocampus segmentation in the 7.0T MR images.

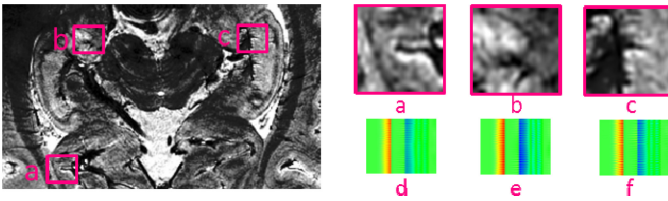


Fig. 2. Demonstration of the moderate power of the hand-crafted features (i.e., Haar features) in representing image patches in a 7.0T MR image

2.1 Learn Hierarchical Feature Representations for Image Patches by ISA

Here, we assume all observed image patches on 7.0T MR image forms a feature space. Then, independent subspace analysis is applied upon image patches, in order to (1) learn basis filters to represent the observed image patches and (2) use the representation coefficients as morphological signature in the feature space to identify the characteristics of each image point in 7.0T image.

Given M basis filters $\mathbf{W} = [w_i]_{i=1, \dots, M}$, we can obtain M responses for each image patch $x^t \in \mathbf{X}$ ($t = 1, \dots, T$) where T denotes the total number of observed image patches. Note that both x^t and w_i are the column vectors. The presence of image feature, i.e., obtained with the inner product $w_i \cdot x^t$, is termed as *response* in this paper. ISA is the unsupervised learning method to learn the basis filters \mathbf{W} from

the observations \mathbf{X} , without requiring independency among all responses $\{w_i \cdot x^t\}$. Instead, ISA seeks for the subspaces in the entire domain of responses and allow dependencies inside each subspace, but independency between any two subspaces. Therefore, similar image features can be grouped into the same subspace to achieve the invariance. To this end, ISA uses the matrix $\mathbf{V} = [v_{ij}]_{i=1,\dots,M,j=1,\dots,Q}$ to represent the structure of Q subspaces, where each entry v_{ij} in \mathbf{V} indicates whether the basis filter w_i is associated with the j -th subspace. \mathbf{V} is usually fixed in training ISA. The objective function of ISA is given by:

$$\widehat{W} = \arg \min_w \sum_{t=1}^T \sum_{j=1}^Q p_j(x^t; W, V), \quad s. t., WW' = I, \quad (1)$$

where p_j is called as the *activation* of particular image patch x^t in the j -th sub-space:

$$p_j(x^t; W, V) = \sqrt{\sum_{i=1}^M v_{ij} (w_i \cdot x^t)^2}. \quad (2)$$

In order to make the training of ISA efficient for high-resolution 7.0T MR images, we follow the method in [15] to construct a stacked two-layered convolutional network, as an extension of ISA by utilizing the technique of stacking and convolution in deep learning. The demonstration of ISA is shown in the bottom of Fig. 3. The input of ISA is the observed image patches, as denoted by boxes. The basis filters and the subspace learned by ISA are shown by triangles and circles, respectively. Specifically, we first extract image patches with a large scale. Then, we follow the sliding window to obtain a set of overlapped image patches but with smaller scale. Since the dimension of these cropped patches are small, we can efficiently train the ISA, thus obtain the activations for all small-scale image patches. Next, we use the combination of the activations from all small-scale patches in the large-scale patch as the input of another ISA in the second layer. The two-layer ISA network is shown in Fig. 3, with blue and red colors denote the 1st and 2nd layer of ISA, respectively. Considering that the dimension of input to the 2nd-layer ISA is still high but redundant, PCA is deployed to reduce the dimension before training of the 2nd-layer ISA. The basis filters learned by the 1st-layer ISA is shown in Fig. 4, where most of them look like the Gabor filters. The result of learned basis filters in the low level is reasonable since edge information is very rich in 7.0T images.

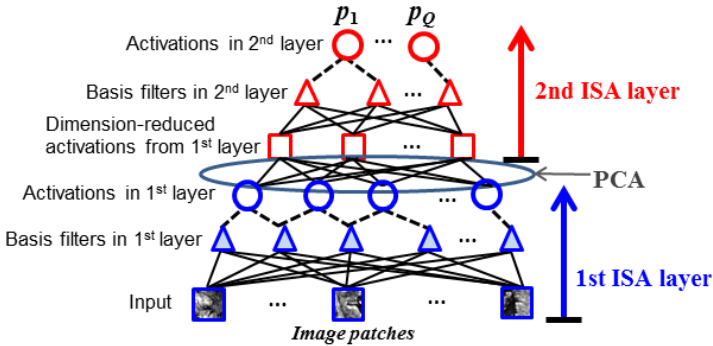


Fig. 3. Stacked convolutional ISA networks for feature extraction

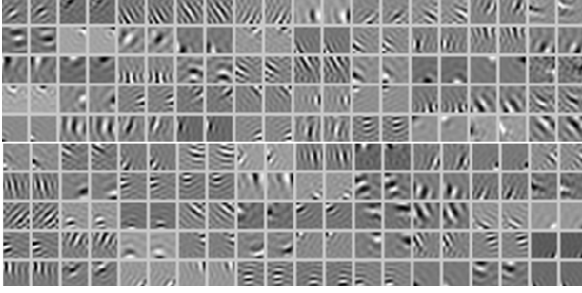


Fig. 4. The learned basis filters by 1st layer ISA

2.2 Segment Hippocampus by Learned Feature Representations

Although 7.0T imaging technique has demonstrated the superior image quality around hippocampus, few automatic segmentation methods have been proposed in the literature. As mentioned in the introduction, the characteristics of 7.0T image, e.g., severe intensity inhomogeneity and high SNR, can degrade the segmentation performance by the existing learning-based methods. Moreover, the alignment of 7.0T images is also not straightforward. Therefore, we extend auto-context model (ACM) [10], which does not require accurate alignment of training images, into a multi-atlases based segmentation framework, as demonstrated in Fig. 5.

In the *training* stage (Fig. 5(a)), totally N images $I = \{I_s(x)|x \in \Omega, s = 1, \dots, N\}$ and their corresponding manual hippocampi labels $L = \{L_s(x)|x \in \Omega, s = 1, \dots, N\}$ are used as atlases. In each atlas space (depicted by dashed boxes in Fig. 5), all other $N - 1$ atlases are linearly aligned onto this atlas and then a number of image features are extracted for each point, within a certain neighborhood. Note that we extract the hierarchical features learned by the stacked convolutional ISA model (as described in Section 2.1), rather than the hand-crafted features.

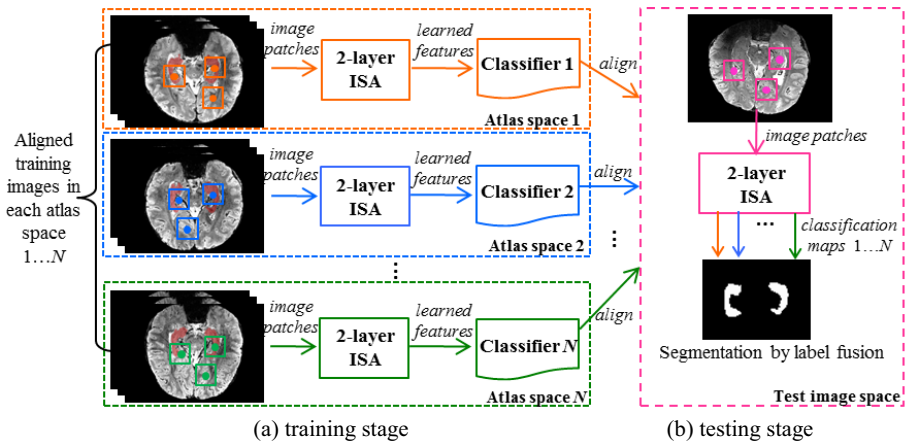


Fig. 5. Schematic illustration of multi-atlases based hippocampus segmentation framework, which consists of training stage (a) and testing stage (b)

Since it is difficult to align 7.0T images, ACM method is deployed in our multi-atlases based segmentation framework. ACM utilize spatial context information, which is iteratively perceived from the probability map of segmented hippocampus at the previous iteration, without requiring the well alignment of training images. After deploying an ACM classifier sequence in each atlas, a set of classifier sequences w.r.t. the number of atlases are trained, where the training samples include not only the underlying atlas but also the linearly aligned other $N - 1$ atlases as mentioned above. In the testing stage (Fig. 5(b)), same image features (in our case, hierarchical features learned by ISA network) are extracted for each subject point. Then, the following steps are repeated for each atlas to predict the label for each subject point: (1) Map the classifiers on each atlas to the underlying subject space by using the affine registration between the atlas and subject image; (2) Predict the probabilistic labeling map by applying the trained ACM classifiers, trained w.r.t. each atlas, to each point in the subject; (3) Fuse all labeling results from all atlases to obtain the final segmentation result.

3 Experimental Results

We demonstrate the performance of hierarchical feature representation through deep learning by incorporating it into the state-of-the-art multi-atlases based segmentation framework. Specifically, we compare our method with the other method in the same multi-atlases framework but incorporating with hand-crafted image features used in [10]. We extract the smaller image patch with the size of $16 \times 16 \times 3$ to train the 1st layer ISA, and the larger image patch the size of $20 \times 20 \times 5$ to train the 2nd layer ISA network, respectively. The initial dimension for patch representation by the 1st ISA layer is 200, while the final dimension by the 2nd layer is 100. We report both qualitative and quantitative segmentation results of hippocampus by our proposed method, with comparison to other method using the hand-crafted features.

In the following experiments, we use totally 20 7.0T MR images, each with the image size $576 \times 576 \times 60$ and voxel resolution $0.35 \times 0.35 \times 0.35mm^3$. A leave-one-out test is used due to the limited number of samples. Specifically, at each leave-one-out case, one image is used as test image, and all other images are used as atlas images. In both training stage and testing stage, the affine registration is used to bring the images to the same space by the flirt algorithm in FSL library.

3.1 Qualitative Results

Fig. 6 shows two typical segmented results by using hand-crafted image features (Fig. 6(b)) and the learned hierarchical feature representations (Fig. 6(c)). It can be observed that the segmented hippocampi with our learned features (Fig. 6(c)) are much closer to the manual ground-truths (Fig. 6(a)), especially for the regions indicated by circles. This result shows the capability of deep learning in extracting the distinctive hierarchical features in 7.0T images for more accurate segmentation.

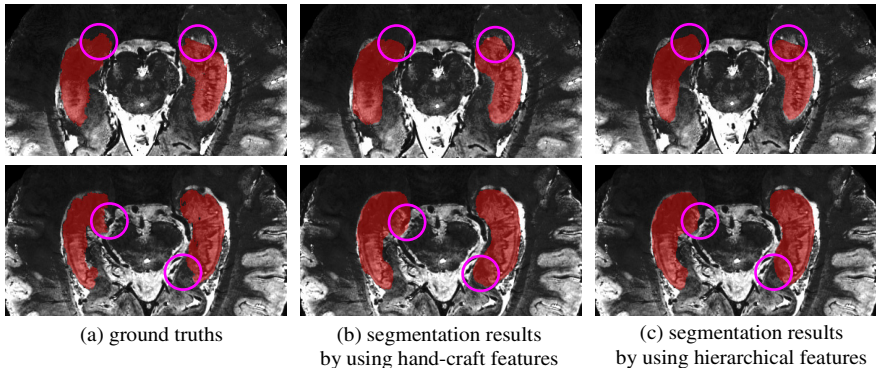


Fig. 6. Comparison of segmented hippocampus regions by (b) using hand-crafted features, and (c) using hierarchical feature representations. Compared to the manual ground truths (a), hippocampus segmentation results with our learned features gives better performance.

3.2 Quantitative Results

For quantitative evaluation of hippocampus segmentation, we use the following 4 overlap metrics: precision (P), recall (R), relative overlap (RO), and similarity index (SI), as defined below:

$$P = \frac{V(A \cap B)}{V(B)}, \quad R = \frac{V(A \cap B)}{V(A)}, \quad RO = \frac{V(A \cap B)}{V(A \cup B)} \quad \text{and} \quad SI = \frac{V(A \cap B)}{\{(V(A) + V(B))/2\}} \quad (3)$$

where $V(A)$ is the volume of ground-truth segmentation such as manual segmentation, and $V(B)$ is the volume of automatic segmentation. Table 1 shows the averaged overlap scores for these 4 above metrics for the 20 leave-one-out cases, indicating that our learned hierarchical feature representations consistently achieve better hippocampus segmentation than hand-crafted features across all 4 metrics.

Table 1. Quantitative comparisons based on the averaged 4 overlap metrics (precision (P), recall (R), relative overlap (RO), and similarity index (SI)) for the 20 leave-one-out cases, which shows the improvements by our method using learned feature representation over other method using hand-craft features in the same segmentation framework (unit: %)

	P	R	RO	SI
By hand-crafted features	84.3	84.7	77.2	86.5
By hierarchical patch representations	88.3	88.1	81.9	89.4

4 Conclusion

In this paper, we proposed using the unsupervised deep learning to extract discriminative image features for segmenting hippocampus in the 7.0T MR images. In view of abundant image details as well as image artifacts, we constructed a two-layer ISA network to seek for the intrinsic basis filters directly from the pre-observed image

patches, and then used the resulting hierarchical feature representations as distinct morphological signature to characterize each image point in the 7.0T MR image. Experimental results demonstrated superior performance of our learned hierarchical features over the hand-crafted image features in substantially improving the segmentation accuracy of hippocampus in 7.0T MR images.

References

1. Zhou, J., Rajapakse, J.C.: Segmentation of subcortical brain structures using fuzzy templates. *NeuroImage* 28(4), 915–924 (2005)
2. Coupé, P., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
3. Khan, A.R., Wang, L., Beg, M.F.: FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using Large Deformation Diffeomorphic Metric Mapping. *NeuroImage* 41(3), 735–746 (2008)
4. Chupin, M., et al.: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation. *NeuroImage* 46(3), 749–761 (2009)
5. van der Lijn, F., et al.: Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage* 43(4), 708–720 (2008)
6. Lötjönen, J.M.P., et al.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage* 49(3), 2352–2365
7. Cho, Z.-H., et al.: Quantitative analysis of the hippocampus using images obtained from 7.0 T MRI. *NeuroImage* 49(3), 2134–2140 (2010)
8. Wang, H., et al.: A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. *NeuroImage* 55(3), 968–985 (2011)
9. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Toga, A.W., Thompson, P.M.: Automatic Subcortical Segmentation Using a Contextual Model. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 194–201. Springer, Heidelberg (2008)
10. Tu, Z., Bai, X.: Auto-Context and Its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1744–1757 (2010)
11. Zhang, S., Zhan, Y., Metaxas, D.N.: Deformable segmentation via sparse representation and dictionary learning. *Medical Image Analysis* 16(7), 1385–1396 (2012)
12. Zhang, S., et al.: Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis* 16(1), 265–277 (2012)
13. Bengio, Y.: Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* 2(1), 1–127 (2009)
14. Shin, H.-C., et al.: Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 99(PrePrints), 1 (2012)
15. Le, Q.V., et al.: Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2011)