PCA-aided Fully Convolutional Networks for Semantic Segmentation of Multi-channel fMRI

Lei Tai[1,3], Haoyang Ye[1,3], Qiong Ye[2] and Ming Liu[3]

Abstract— Semantic segmentation of functional magnetic resonance imaging (fMRI) makes great sense for pathology diagnosis and decision system of medical robots. The multi-channel fMRI provides more information of the pathological features. But the increased amount of data causes complexity in feature detections. This paper proposes a principal component analysis (PCA)-aided fully convolutional network to particularly deal with multi-channel fMRI. We transfer the learned weights of contemporary classification networks to the segmentation task by fine-tuning. The results of the convolutional network are compared with various methods e.g. k-NN. A new labeling strategy is proposed to solve the semantic segmentation problem with unclear boundaries. Even with a small-sized training dataset, the test results demonstrate that our model outperforms other pathological feature detection methods. Besides, its forward inference only takes 90 milliseconds for a single set of fMRI data. To our knowledge, this is the first time to realize pixel-wise labeling of multi-channel magnetic resonance image using FCN.
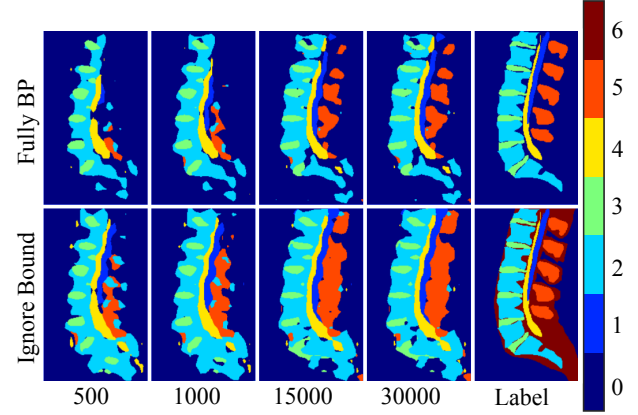
Fig. 1. Representation of number: 0-back ground, 1-cerebrospinal, 2-vertebral bodyfluid, 3-lumbar disc, 4-spinal fluid, 5-bones, 6-ignore area. Different test result after 500, 1000, 15000, 30000 training iterations and the label instruction of two labeling strategies.

I. INTRODUCTION

Medical-image analysis is indispensable in modern computer-aided diagnosis, therapy planning, and execution. Management of the related pathology relies on several different kinds of imaging processing technologies including positron emission tomography (PET), X-ray computer tomography (CT), ultrasound and magnetic resonance imaging (fMRI) [1]. MRI is amongst the most useful technique for visualization. MRI shows high contrast and high resolution for different body tissues in a non-invasive way. Semantic segmentation of magnetic resonance image is an essential step for diagnosis with the abstraction of relevant features. Through segmentation in the feature space, semantic segmentation minimizes the computational complexity in fMRI applications [2]. For both diagnosis and treatment strategies, accurate classification and segmentation of the tissues are necessary and primordial, such as the quantitative volume measurement of different brain structures in brain fMRI segmentation [3].

Convolutional networks have been applied to a variety of recognition and perception tasks recently, e.g. image classi-

fication [4], object detection [5] and semantic segmentation [6]. The great success of Convolutional Neural Network (CNN) is attributed to the outstanding performance of trained hierarchical structure in data representations. The weights of CNN are typically calculated by back-propagation and gradient descent. This self-driven method has been keeping showing its unlimited potential in other conventional robotic fields [7] [8] as well.

Pixel labeling, which is also called semantic segmentation, is much more challenging compared with image recognition and object detection. The scale of perception changes from a whole image to pixel-wise elements. For image classification, the coarse distribution is taken as the output. But for semantic segmentation, the output is a dense prediction at every pixel location. Not like the traditional recognition and segmentation methods [9], [10], [11], where mutual information is used manually, fully convolutional networks (FCN) [6] is a CNN variant. It achieves the state-of-art effect in semantic segmentation task for most of the benchmarks.

However, not like computer vision tasks with various datasets such as ImageNet [12], it is hard to collect an fMRI dataset for specific tissues of the human body and specific pathology features. Regular semantic segmentation dataset is comprised of 3-channel RGB images, while the parameters of fMRI are not constant based on the application and usage. Realizing such a segmentation in a small dataset with multi-channel fMRI means massive work load for both supervised expertise and medical diagnosis. The multi-

[1]Lei Tai and Haoyang Ye are with MBE, City University of Hong Kong. {lei.tai, hy.ye}@my.cityu.edu.hk

[2]Qiong Ye is with Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Zhejiang, China. 94301699@qq.com

[3]Lei Tai, Haoyang Ye and Ming Liu are with Department of ECE, the Hong Kong University of Science and Technology. eelium@ust.hk

channel infrastructure brings more co-related information, but also increases the data complexity. Training from small datasets like particular pathology and disease features can help build an expert system to recognize the same feature in the scene, which will bring great convenience for doctors. Another obstacle for pixel-wise segmentation in fMRI is the unclear boundary. Not like the typical RGB images by digital cameras, junction area between different tissues in fMRI are hardly recognizable even for experienced doctors. It also adds the complexity of the labeling process.

Motivated by the facts mentioned above, we stress the following contributions and features of this work:

- We propose a PCA-aided FCN structure to process 31-channel fMRI and to achieve pixel-wise semantic segmentation for primary tissues. Through PCA, the dimension of the origin data will be reduced to three and be further taken as the input of the fully convolutional network.
- The fine-tuning is implemented on a small-sized dataset as transfer learning. The experiment results show outperforming efficiency compared with conventional methods for fMRI feature learning.
- A revised labeling strategy is proposed, which ignores the junction areas between tissues, which significantly decreases the complexity of semantic labeling for segmentation.

The rest of paper is organized as follows. We present related works in fMRI and pixel labeling in Section II. In Section III, we describe our pre-processing configurations of the dataset with PCA and normalization. The configuration of the whole FCN structure is also introduced in this section. The detail of the training and evaluation on specific datasets are then presented in Section IV. At the end, Section V concludes the paper.

II. RELATED WORK

A. fMRI Segmentation

Most of conventional brain fMRI segmentation tasks are probabilistic, based on clustering methods like *k-means*. Usually, the MR images are represented as a collection $X = \{x_1, x_2, \cdots, x_n\}$ with every element representing a pixel of the image. The clustering algorithm returns $C = \{c_1, c_2, \cdots, c_m\}$ as the cluster centers. C are commonly initialized in random.

Barrah et al. [1] used a revised *fuzzy clustering* algorithm to segment fMRI. Standard fuzzy c-means considers the clustering an optimization problem. The objective function is defined as:

$$\arg \min_C \sum_i^n \sum_j^m w_{ij}^h \|x_i - c_j\|^2$$

where w_{ij} is the degree to which element x_i belongs to cluster c_j and h is the fuzzifier to determine the level of cluster fuzziness. Euclidean distance is used to evaluate the element and cluster center directly. After that, w_{ij} and c_j

can be estimated iteratively, as

$$w_{ij} = \frac{1}{\sum_{k=1}^m \left(\|x_i - c_j\| / \|x_i - c_k\| \right)^{\frac{2}{m-1}}}$$

$$c_i = \sum_{j=1}^m w_{ij}^h x_j / \sum_{j=1}^m w_{ij}^h$$

In [1], the nearby area of the element in an image was considered as well. *Gaussian mixture model* was also used to address brain MR images clustering [3]. Compared with *fuzzy clusterng* mentioned above, the Gaussian kernel takes the place of norm functions between elements and cluster centers.

There are some challenges about ventricle fMRI segmentation, such as Statistical Atlases and Computational Modeling of the Heart (STACOM) [13] and Medical Image Computing and Computer Assisted Intervention (MICCAI) [14]. Petitjean supplied a review of conventional methods for left ventricle segmentation[15], which consisted of graph cuts and deformable models, etc.

All of these methods provided a reasonable segmentation result for fMRI datasets. However, these statistical-based or feature-based method were limited, because they tend to overfit on specific dataset like ventricle. Their utility in a generic dataset was extremely limited as well.

B. ConvNet in Semantic Segmentation

The convolutional neural network is leading the artificial intelligence right now. Related research [16] just beat the best human player in Go game which was once regarded as the most difficult task for artificial intelligence. In image classification, deep residual network [17] with hundreds of convolutional layers drove the network to learn identity feature and improve the accuracy of classification tasks in ImageNet [12]. Faster-rcnn [5] addressed all of the tasks in object detection including region proposal, bounding box regression, instance location, and classification. CNN related methods take the state-of-the-art place in almost every computer vision tasks. Not only in vision fields, the successes of RNN and LSTM in natural language processing [18] and speech recognition [19] are all based on revised convolutional neural network. In terms of robot control, motivated by the game controller developed by Google [20], deep reinforcement learning is implemented to teach the robot be adapted to the unfamiliar environment and achieve obstacle avoidance ability [7] [8]. We adopt most of these concepts in this work.

Among various vision tasks, a high-level prediction like semantic segmentation is much more complicated compared with simply predicting image context as single output. While for pixel-wise classification, the computation source is thousands times more than the image classification.

Long *et al.* [6] re-architected and fine-tuned classification nets to achieve dense prediction of semantic segmentation directly. They improve the efficiency of convets in semantic segmentation dramatically because FCN extended a

convolutional network to arbitrary-sized inputs [21]. Deep Convolutional Neural Networks (DCNN) were not sufficient to achieve semantic segmentation separately because of the poor localization property. Chen [22] combined the feature map at the final layer of DCNNs with a fully connected Conditional Random Field (CRF). The main breakthrough in [22] was the intersection of convolutional network and probabilistic graphical models. Notice that, CRF and FCN are the state-of-the-art semantic segmentation methods nowadays. [23] applied FCN processing cardiac segmentation in the ventricle dataset mentioned in Section II-A, which also focused on a unique dataset but not generic multi-channel fMRI.

III. PCA-AIDED FULLY CONVOLUTIONAL NETWORKS

Our approach benefits a lot from the recent popular deep net-related works including image classification [4], semantic segmentation [6] and transfer learning [24]. After PCA, the processed fMRI data are taken as the input of FCN.

A. Data Gathering and Preparation

This study was approved by the local institutional ethic review board. Written informed consent was obtained from all volunteers. Six young and healthy volunteers (age range, 40 years; median age, 38 years; three males and three females) were recruited in this study. In addition to standard criteria for exclusion from fMRI, further exclusion criteria were applied using the absence of cervical spine related pain and disorders.

The dataset focuses on the vertebral area of human being. Except the background, there are five primary tissues including bones, spinal fluid, lumbar disc, vertebral body and cerebrospinal fluid shown in Fig. 1.

fMRI examinations were carried out on 3T clinical MR scanner (Achieva, Philips Healthcare, Best, The Netherlands) with standard abdomen coil. Mid-sagittal multi-echo T2-weighted images were acquired using a multi-shot spin-echo sequence with echo-planar readout and following parameters shown in Table I. Here *TE* means echo time. The intensity of image under different echo time reflects the attenuation of tissues. With 32 different echo times, 32 channels representing the different intensity of various tissues. The first echo of the sequence was excluded from further analysis to avoid the effect of the stimulated echo [25]. After cutting the black background edge, the dimension of every sample in the dataset turns to be $31 \times 256 \times 154$.

B. PCA Processing

It is evident that the 31-channel fMRI provide more information in detail about the pathology features. But because 31-channel images increase the calculation complexity dramatically, a dimension reduction method is necessary. Besides, state-of-the-art CNN structures were usually implemented on vision tasks which take raw RGB images with three channels as input. Preliminaries experiments also tell that if we attach the images with 31 channels as input layer

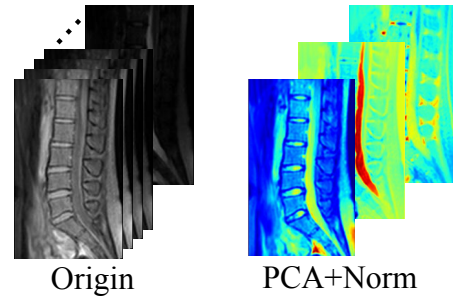| Parameter	Value	unit
FOV	$240 \times 155 \times 5 (FH \times AP \times RL)$	mm^3
matrix	120×72	-
reconstruction	32	-
slice thickness	5.0	mm
No. of slice	1	-
No. of TEs	32	-
TE/TR	$8.0 \times 32/1202$	ms
NA	8	-
fat suppression	SPIR	-
scan time	235.5	s

Origin   PCA+Norm

Fig. 2. Data transformation from 31 channels to 3 channels after PCA processing. Origin data is collected from 6 volunteers. The dataset shows vertebral area of human beings.

to the network, the weights of the network is very difficult to converge. These difficulties are to be moderated.

In PCA analysis, we vectorize every channel of the sample to a vector. The sample is transformed to a matrix X as 31×39424. PCA is a variant matrix transformation as

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X w_{(s)} w_{(s)}^T$$

Here $w_{(s)}$ means the related eigenvector of XX^T's sorted eigenvalues. And $w_{(s)}$ is also a united vector. It can be calculated like

$$w_{(1)} = \arg\max \left(\frac{w^T X^T X w}{w^T w} \right)$$

The left part of Fig. 2 shows a sample of the origin 31-channel images and the transformation result after PCA and normalization processing. Fig. 3 shows all of the 31 singular values of X which is also the square root of eigenvalues of XX^T from all of the six samples of the datasets. The first three singular values take the weights of all the singular values in more than 99.9%. That means the transformation is more efficient and meaningful. The largest three singular values of X can almost represent all of the origin information in the raw sample equivilently.

The fMRI datasets have a broad range of pixel intensities. The different ranges significantly influence the accuracy of segmentation. Even after PCA processing, the ranges of the three channels are still quite irregular. To locate the maximal and minimal elements among the three channels, we normalize the three channels to a range of (0,255),
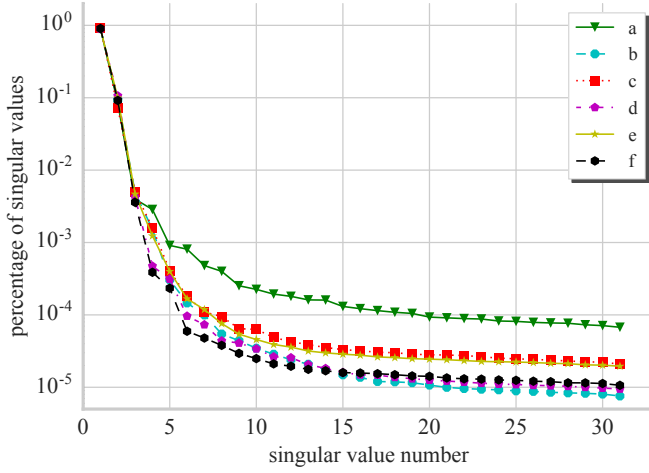
Fig. 3. Percentage of singular values of all the samples in datasets. Singular values are sorted and the percentage is shown in *log* scale.

which is the range of 8-bit images. The right part of Fig. 2 illustrate the sample condition after PCA and normalization. The supervising learning procedure will be introduced later in Section III-C. The labeling process is tackled on the origin sample before PCA pre-processing. We propose two different labeling strategies as shown in Fig. 1, dealing with five categories of primary tissues in the vertebral fMRI datasets, as follows:

- *Fully-BP:* The origin image is aggressively segmented to six parts including all of the five primary classes as introduced in Section III-A and the background area. The predictions in every pixel location are to be added in the loss function and used to execute back propagation for weights.
- *Ignore-Bound:* Among the primary tissues, the boundary (separation plane) is vague even for human inspectors. We ignore these boundary parts and label no targets for pixels located here. The forward predictions of these parts will not be considered when calculating the loss. Naturally, there is no back propagation through these locations.

C. FCN Implementation

Fig. 4 illustrates our proposed FCN architecture. In the convolutional networks, each layer is a three-dimensional array of size $h \times w \times d$, where d is the channel dimension and h, w are the spatial sizes. Here the dimension of the input data is following the pre-processed fMRI data. The size of input array is thus $256 \times 154 \times 31$.

1) Dense Prediction: The feature learning of FCN uses the structure of *VGG-16* net [4] which won ILSVRC14. As introduced in [6], all of the fully connected layers of *VGG-16* are converted to convolutions. Finally, there are 15 convolutional (Conv) layers for feature learning. To increase non-linearity, every Conv layer is followed by a Rectified Linear Unit (ReLU) activation function layer. They are represented as red cubes in Fig. 4. The number under the Conv+ReLU

cubs in Fig. 4 is the number of channels of the output data related to this cube. A 1×1 convolution layer with channel dimension six is appended to predict scores for each of the tissue classes including the background. As shown in Fig. 4, actually the Prediction cube is also a convolution layer. All of the output channel numbers of prediction convolution layers is six as well. The Deconv layer will upsample the coarse output to the similar size as input (256×154).

2) Hierarchy Combination: Max pooling layers, represented by blue cubes, are implemented after several Conv+ReLU cubes to spatially reduce the data size. We use the same strategy to combine layers of the feature hierarchy as introduced in [6]. The output of *Pool3*, *Pool4*, *Pool5* are to be unsampled to with different strides. As the sequence illustrated in Fig. 4, *Deconv1* and *Predict2* will be firstly added conjugatedly and unsampled to *Deconv2*. Then *Deconv2* and *Predict3* will be added in sequence and unsampled to *Deconv3*. The dimension of *Deconv3* is $6 \times 256 \times 154$. At last, a Softmax layer is appended to output the final prediction.

IV. EXPERIMENTS AND RESULTS

A. Training and Test

The whole structure is trained end-to-end and the initial weights of network are transferred from the training result of VOC dataset [26] in *voc-fcn8s* [1] structure which was pre-trained with the parameters in [6]. In the fine-tuning procedure, we use Stochastic Gradient Descent (SGD) method with momentum. The latest research [27] showed that setting the batch size to one is beneficial for semantic segmentation. We tried different learning rate from 10^{-5} to 10^{-14}, and smallest learning rate results in the best performance. One possible explanation is that the initialization from ImageNet-trained weights is important, so the tuning that makes it adaptive to the new datasets should be tiny and slow. Training parameters is described in detail in II.

We use Caffe [28] *python* interface as the deep learning framework to train and test all of the data and models. A unit NVIDIA Tesla K40c is used to support the computation.

There are six MR images in the vertebral datasets as illustrated in Fig. 3. We separate one of them as the test sample. Fig. 5 shows the training and test loss decreasing with the iteration number increasing. The two strategies introduced in Section III-B are all implemented and tested. All of the losses experience substantial reduction at first 5000 steps. After that, because of the existence of ignored area, the loss of *Ignore-Bound* strategy is obviously less than the one

---

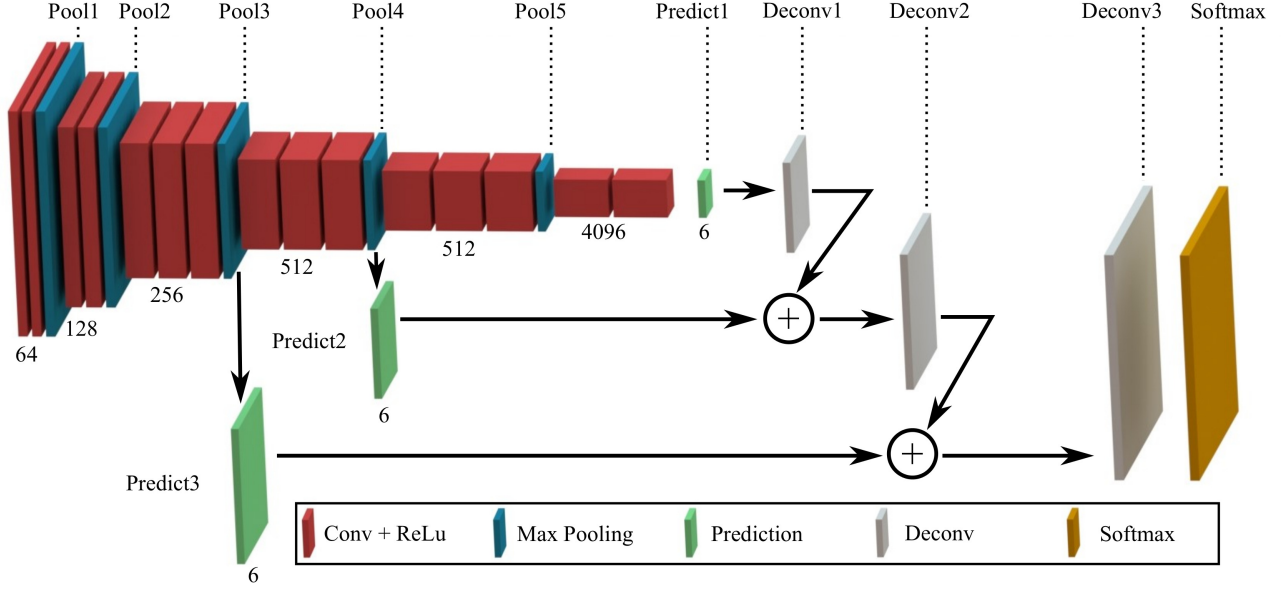[1] http://dl.caffe.berkeleyvision.org/fcn8s-heavy-pascal.caffemodel

Fig. 4. The structure of our proposed fully convolutional neural network. Feature learning part use the structure of *VGG-16* net[4]. After feature learning, the pooling output of different stages are tackled with upsampling and merged. The final layer uses a Softmax layer to output dense classification result matrix.
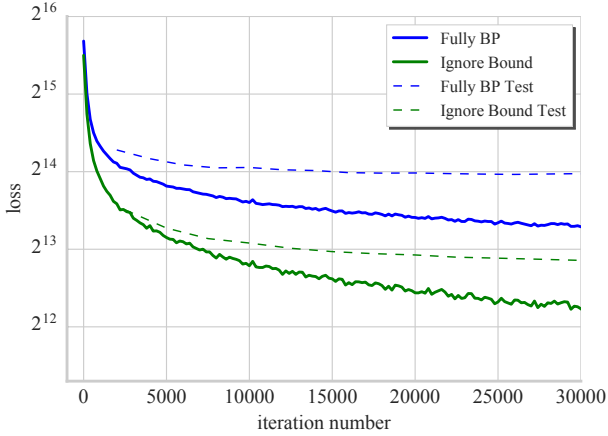

Fig. 5. The decrease process of training and test loss of two different label strategies. The multi-nominal logistic loss of all of the pixels in the sample image are added.

with the *Fully-BP* strategy. The test losses of both strategies are converged stably after 15000 iterations.

B. Results

A naive K-Nearest-Neighbor (KNN) method, which is commonly used nowadays, is implemented as comparison as well.

In this experiment, there are five samples in the training dataset as mentioned in Section III-A. From the training dataset, we got $5 \times 256 \times 154$ pixels for each scan. After PCA and normalization process, there is a 3-dimensions vector for every pixel location. We calculate the median 3-dimensional vectors for pixels belong to same class and denote it by y_j for class j. For all of the six classes introduced in Section III-A, let $Y = \{y_1, y_2, \ldots y_6\}$ be the medians set for all of six classes. For a pixel in the test sample after PCA and normalization process, let the 3-dimensional pixel vector be x. We can locate the least normalized inner product between x and the element in the median sets Y as:

$$\arg\min_j \frac{x^T y_j}{\|x\|\|y_j\|}$$

Then j is the prediction result of pixel x. By scanning the whole image, we can get the prediction for every pixel in the test data.

TABLE III
PARAMETER REPRESENTATIONS FOR CRITERIA

Parameters	Definitions
i, j	specific class number
n_{ij}	pixels of class i predicted to belong to class j
n_c	number of classes (6)
s_i	pixels labeled with class i ($\sum_j n_{ij}$)

We use several common semantic segmentation and scene parsing methods to compare the experiment results, using criteria like pixel accuracy and region intersection over union (IU). Based on the parameters introduced in Table III, we compute:

- mean IU:

$$(1/n_c) \sum_i n_{ii}/(s_i + \sum_j n_{ji} - n_{ii})$$

- frequncy weighted IU (fw IU):

$$\left(\sum_k s_k\right)^{-1} \sum_i s_i n_{ii}/(s_i + \sum_j n_{ji} - n_{ii})$$
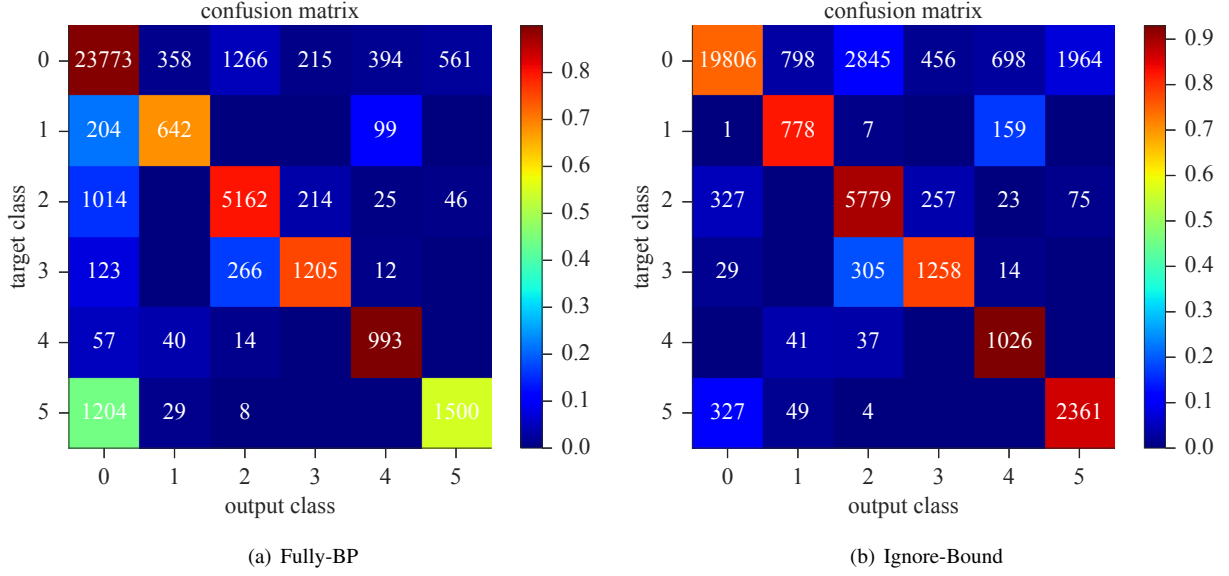
Fig. 6. The confusion matrix of test results with two strategies after 15000 iterations of training. Representation of number: 0-back ground, 1-cerebrospinal, 2-vertebral bodyfluid, 3-lumbar disc, 4-spinal fluid, 5-bones. The number in the grid located at row i and column j means pixels in the test sample labeled with class i and predicted to belong to class j, which is n_{ij} introduced in Table III. The color of this grid represents the precision (n_{ij}/s_i). The number of $n_i i$ on the diagonal is also called true positive count.

TABLE IV

TEST RESULTS FOR FCN BASED ON TWO STRATEGIES AT THE RELATED TRAINING SECTION AND THE NAIVE METHOD.

Methods	all classes				main tissues			
	mean IU	fw IU	pixel acc	mean acc	mean IU	fw IU	pixel acc	mean acc
Naive KNN	13.1	13.0	23.0	44.6	13.5	16.4	44.5	38.3
Fully-BP 10000	58.8	73.3	83.9	75.8	54.4	57.3	74.1	80.1
Fully-BP 15000	59.6	73.8	84.4	76.2	55.2	58.0	73.9	80.5
Fully-BP 30000	**60.7**	**74.4**	**84.7**	78.3	**56.5**	**59.3**	75.1	81.7
Ignore-Bound 10000	54.8	66.5	78.4	**84.2**	54.4	54.5	87.1	**83.8**
Ignore-Bound 15000	55.0	66.9	78.7	84.0	51.5	54.9	87.1	83.5
Ignore-Bound 30000	54.1	66.3	78.2	83.8	50.6	54.6	**87.6**	82.8

- pixel accuracy:

$$\sum_i n_{ii} / \sum_i s_i$$

- mean accuracy:

$$(1/n_c) \sum_i n_{ii}/s_i$$

For the particular target of segmentation for medical images, the background class is not as essential as the other meaningful tissues. When we calculate the loss in the back propagation procedure of training, the same metrics for all of the five primary tissues except the background are also introduced as shown in II.

The evolution of test results and label maps are described in Fig. 1. As the result of loss reduction, the shapes of the segmentation results progressively tended to the same as the ground-truth at the first several thousands steps. After 15000 iterations, there is hardly any revision anymore. By comparison, we find that the *Fully-BP* strategy prefers to predict the junction area as background. On the other hand, the *Ignore-Bound* strategy is prone to regard the junction area

as tissues especially for bone area. The confusion matrices of the test result after 15000 iterations for both strategies which reflect the difficulties of prediction for bones are illustrated in Fig. 6. An probable explanation is that the bone area is the most difficult one to label with unclear boundary mostly as shown in Fig. 2. There may be too many labeling mistakes in this area with the *Fully-BP* strategy. For the precision of different classes, the *Ignore-Bound* strategy substantially improves the precision of all classes except the background part. However, it takes the cost to increase the false positive counts for primary tissues as well.

Comparison among various methods and FCN methods against different training periods are listed in Table II. FCN methods show great improvement compared with other methods. Another fact is that FCN only needs 90 milliseconds to calculate the forward prediction, and the conventional method takes 1.4 seconds to finish the traversal analyzing for every pixel in the whole image. From Table II, the *Fully-BP* strategy shows high performance in mean-IU and frequency-weighted-IU. The *Ignore-Bound* strategy has increased accuracy especially for primary tissues. For the five primary classes of tissues except the background, the differences in

mean IU and fw IU are also reduced compared with the *Fully-BP* strategy. And the best performance of the *Ignore-Bound* strategy appeared after around 10000 iterations. It takes less time to achieve the satisfied result compared with the *Fully-BP* strategy which need 30000 iterations. Generally speaking, the *Ignore-BP* strategy shows benefits in converging speed and accuracy.

V. CONCLUSION

This papers demonstrated the effectiveness to apply FCN deep-learning methods on semantic segmentation of multi-channel fMRI. The utility of a PCA-aided fully convolutional network model is proved under efficient end-to-end training using a small-sized dataset. It indicates that the transfer learning and *Ignore-Bound* label strategy are evidentially sufficient when collecting small datasets, labeling them and realizing the pixel-wise classification in near real-time. The proposed framework outperforms other state-of-the-art methods in precision.

A few limitations to our approach remain. Most notably, as with most CNN approaches, the training parameters is decided after a lot of attempts. Fine-tuning from other pre-trained models is also time consuming. The labeling process of *Ignore-Bound* strategy need an appropriate threshold to distinguish the junction area.

In the future, the labeling strategy can be further improved considering the limitations in parts of the experimental results. The combination with CRF introduced in [22] should also be implemented and tested. Eventually, we will not only test on the tissue segmentation but also apply the segmentation method on pathology related areas directly.

REFERENCES

[1] H. Barrah, A. Cherkaoui, and D. Sarsri, "Agent based segmentation of the mri brain using a robust c-means algorithm," *Journal of Computer and Communications*, vol. 4, no. 10, p. 13, 2016.

[2] L. Clarke, R. Velthuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, and M. Silbiger, "Mri segmentation: methods and applications," *Magnetic resonance imaging*, vol. 13, no. 3, pp. 343–368, 1995.

[3] Y. Huang, J. P. Dmochowski, Y. Su, A. Datta, C. Rorden, and L. C. Parra, "Automated mri segmentation for individualized modeling of current flow in the human head," *Journal of neural engineering*, vol. 10, no. 6, p. 066004, 2013.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[7] L. Tai, S. Li, and M. Liu, "A deep-network solution towards model-less obstacle avoidence," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2016.(IROS 2016)*, 2016.

[8] L. Tai and M. Liu, "A robot exploration strategy based on q-learning network," in *Real-time Computing and Robotics (RCAR) 2016 IEEE International Conference on*, Angkor Wat, Cambodia, June 2016.

[9] M. Liu, L. Wang, and R. Siegwart, "Dp-fusion: A generic framework for online multi sensor recognition," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on*. IEEE, 2012, pp. 7–12.

[10] M. Liu, F. Colas, and R. Siegwart, "Regional topological segmentation based on mutual information graphs," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3269–3274.

[11] M. Liu, F. Colas, F. Pomerleau, and R. Siegwart, "A markov semi-supervised clustering approach and its application in topological map extraction," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 4743–4748.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[13] A. Suinesiaputra, B. R. Cowan, A. O. Al-Agamy, M. A. Elattar, N. Ayache, A. S. Fahmy, A. M. Khalifa, P. Medrano-Gracia, M.-P. Jolly, A. H. Kadish, *et al.*, "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images," *Medical image analysis*, vol. 18, no. 1, pp. 50–62, 2014.

[14] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[15] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical image analysis*, vol. 15, no. 2, pp. 169–184, 2011.

[16] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[18] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[19] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition," in *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*. IEEE, 2012, pp. 4277–4280.

[20] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[21] O. Matan, C. J. Burges, Y. LeCun, and J. S. Denker, "Multi-digit recognition using a space displacement neural network," in *NIPS*, 1991, pp. 488–495.

[22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[23] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," *arXiv preprint arXiv:1604.00494*, 2016.

[24] Y. Bengio *et al.*, "Deep learning of representations for unsupervised and transfer learning." *ICML Unsupervised and Transfer Learning*, vol. 27, pp. 17–36, 2012.

[25] H. Takashima, T. Takebayashi, M. Yoshimoto, Y. Terashima, H. Tsuda, K. Ida, and T. Yamashita, "Correlation between t2 relaxation time and intervertebral disk degeneration," *Skeletal radiology*, vol. 41, no. 2, pp. 163–167, 2012.

[26] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[27] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1605.06211, 2016. [Online]. Available: http://arxiv.org/abs/1605.06211

[28] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.