GENERATION OF SYNTHETIC STRUCTURAL MAGNETIC RESONANCE IMAGES FOR DEEP LEARNING PRE-TRAINING

Eduardo Castro⋆ Alvaro Ulloa† Sergey M. Plis⋆ Jessica A. Turner† Vince D. Calhoun†

⋆The Mind Research Network, Albuquerque, NM ⌂Georgia State University, Atlanta, GA
†The University of New Mexico, Albuquerque, NM

ABSTRACT

Deep learning methods have significantly improved classification accuracy in different areas such as speech, object and text recognition. However, this field has only began to be explored in the brain imaging field, which differs from other fields in terms of the amount of data available, its data dimensionality and other factors. This paper proposes a methodology to generate an extensive synthetic structural magnetic resonance imaging (sMRI) dataset to be used at the pre-training stage of a shallow network model to address the issue of having a limited amount of data available. Our results show that by extending our dataset using 5,000 synthetic sMRI volumes for pre-training, which accounts to approximately 10 times the size of the original dataset, we can obtain a 5% average improvement on classification results compared to the regular approach on a schizophrenia dataset. While the use of synthetic sMRI data for pre-training has only been tested on a shallow network, this can be readily applied to deeper networks.

Index Terms— structural MRI, simulation, deep learning, pre-training, schizophrenia.

1. INTRODUCTION

Deep learning methods have been proven successful on very diverse application areas such as speech, object and text recognition, natural language processing, information retrieval and many others [1]. For supervised learning tasks, deep learning is usually preceded by an unsupervised pre-training phase which seems to play an important role as a regularization mechanism, in addition to providing a better parameter initialization for supervised learning [2, 3].

While pre-training has been shown to aid supervised learning, it is also true that a fair amount of training data is needed to obtain reasonable parameter estimates since the training criterion is non-convex and involves multiple local minima [2]. In fact, the dimensionality of the parameter space scales linearly with the dimensionality of the analyzed data. These are some of the many challenges of applying deep learning to magnetic resonance imaging (MRI) as these datasets are usually composed of at most a few hundred subjects brain volumes, which in turn span several hundreds of thousands of volumetric elements (voxels).

In spite of the challenges of deep learning on neuroimaging, its application to MRI seems promising, as prior work has shown that a restricted Boltzmann machine (RBM) can detect spatial components estimated by independent component analysis (ICA) on functional MRI (fMRI) data [4]. Furthermore, Plis et al. [5] extended this work by showing that more complex (deep) models are capable of better characterizing mental illness by analyzing structural MRI (sMRI) data. These findings are remarkable and prove that deep learning has the potential to address several challenges of brain imaging research. Nevertheless, new strategies are required to better adapt deep learning to MRI data processing.

This paper attempts to improve the classification accuracy of schizophrenia reported in [5] by using a shallow network model along with a novel approach to resample sMRI data based on its ICA decomposition. This resampling approach allows us to arbitrarily increase the amount of structural data to be used at the pre-training stage of this shallow model, which addresses the issue of having a limited amount of data available. By circumventing this problem, we would be capable of finding better parameter estimates of our model at the pre-training stage, which in turn should render a better feature representation at the model hidden layer as well as better classification results.

2. MATERIALS AND METHODS

2.1. Dataset

2.1.1. Participants

We use sMRI data from four schizophrenia studies conducted at Johns Hopkins University, the Maryland Psychiatric Research Center, the Institute of Psychiatry (IOP), London, UK, and the Western Psychiatric Research Institute and Clinic at the University of Pittsburgh, which was used in [6]. This dataset is composed of 198 schizophrenia patients (121 males; age = 39.68±12.12, range 17–81) and 191 healthy controls (97 males; age = 40.26±15.02, range 16–79) matched for age, gender and scanner location.

2.1.2. MRI Parameters

MRI volumes were obtained on a 1.5 T Signa GE scanner on all sites. Almost identical parameters were used at each site (35 ms TR, 5 ms TE, 45 degrees flip angle, 1 excitation, 1.5 mm slice thickness, 24 cm field of view, and a matrix size of 256 × 256), except for IOP data, which was obtained using a 35 degree flip angle and a 20 cm field of view.

2.1.3. Pre-processing

The obtained $T_1$-weighted images were preprocessed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/software/spm5/). These images were normalized to the MNI space, interpolated to voxel dimensions of 1.5 mm$^3$ and segmented into gray matter, white matter and cerebrospinal fluid images, using unified segmentation [7]. The resulting gray matter concentration (GMC) images were then smoothed with an isotropic 8 mm full width at half maximum Gaussian filter.

The correlations of the images with the mean GMC image across subjects were estimated and standardized, regarding those images...
where \( z > 2 \) as noisy. Fifty four images were discarded based on this analysis, yielding a final dataset composed of 335 GMC images. These images, which are composed of 60465 voxels, were input to a fully connected shallow network.

### 2.2. Shallow Network Model

RBMs are probabilistic graphical models that can be interpreted as stochastic neural networks that learn a probability distribution over its set of inputs [8]. They can be modeled by using a two-layer network of visible units, which correspond to the input data features (in our case GMC at each voxel), and hidden units, which learn underlying explanatory data features. Probability distributions over hidden and/or visible units are defined in terms of the configuration energy function, which for real data that is approximately normally distributed can be defined as

\[
E(v, h) = - \sum_{ij} \frac{v_i}{\sigma_i} W_{ij} h_j - \sum_i \frac{(a_i - v_i)^2}{\sigma_i^2} - \sum_j b_j h_j, \quad (1)
\]

where \( W_{ij} \) is the weight between visible unit \( v_i \) and hidden unit \( h_j \), \( a_i \) and \( b_j \) are their associated biases and \( \sigma_i \) is the standard deviation associated to \( v_i \). The parameters learned by an RBM are the weights \( W \) and the biases \( a \) and \( b \). In general, parameters \( \sigma_i \) have to be learned with the other parameters. However, normalizing the distribution of each feature to have zero mean and unit variance and setting \( \sigma_i = 1 \) works well and is faster in practice [9].

The weights associated to a hidden unit are referred to as the receptive field of that hidden unit and provide an interpretation of the data representations learned by an RBM. By incorporating an \( l_1 \)-norm regularization term for \( W \), an RBM applied to MRI data tends to learn more robust representations. In addition, the receptive fields of the hidden units tend to be clustered on specific regions of the brain, providing a better interpretation of the data.

By incorporating a regularization term on the parameters \( W \) we address to a certain extent the possibility of overfitting our limited training data. Nevertheless, we incorporate another criterion, unit dropout [10], to further regularize our model. Succinctly, this technique temporarily removes a given unit of the network with a fixed probability for each data presentation.

Training an RBM accounts to the unsupervised pre-training stage of our supervised model. After pre-training, the model is treated as a feed-forward neural network and trained (fine-tuned) via the error backpropagation algorithm. To do so, the network is augmented by adding a softmax function layer on top of the RBM hidden layer.

### 2.3. Independent Component Analysis

ICA is a matrix factorization technique in which the observed data matrix \([X]_{n \times m}\) is factorized into a mixing matrix \([A]_{n \times c}\) and a source matrix \([S]_{c \times m}\), where \( n \) is the number of samples, \( m \) is the number of measured variables, and \( c \) is the number of sources. This matrix factorization is possible given that the sources are mutually independent and non-Gaussian. The infomax algorithm [11] attempts to estimate the true sources \( S \) by maximizing the joint entropy of the estimated sources \( \hat{S} = \hat{U}X \) as given by

\[
\hat{S} = \text{argmax}_S \{ H(\hat{S}) \},
\]

where \( H(\cdot) \) denotes the joint entropy function and \( \hat{U} \) is the pseudo-inverse of \( A \), which is referred to as the unmixing matrix.

<fig.1> Synthetic sMRI data generation flow diagram.

### 2.4. Generation of sMRI Synthetic Data

We propose a method for the generation of synthetic GMC images that exploits the ICA decomposition of real data to emulate spatial and subject variability.

When factorizing GMC data organized into a subject by voxel matrix, the mixing matrix \( A \) represents the subjects’ loading patterns, i.e., how each source is weighted across subjects. On the other hand, the rows of \( S \) represent the sources, which are weighted patterns of voxels. The number of sources can be estimated using the criterion defined in [12].

The proposed method builds from two main assumptions: (1) Given a representative sample of the population, the estimation of sources from the observed data will be a good approximation of the true sources. (2) A row set of \( A \), corresponding to individuals of a given diagnostic group (controls or patients), share statistical properties inherent to the population of subjects on this group.

First, we de-mean the SMRI data in the voxel dimension to factorize it using ICA. Then, we divide the estimated loading matrix \( A \) in two sub-matrices \( A_{HC} \) and \( A_{SZ} \), corresponding to the healthy control and schizophrenia patient groups respectively. Next, our method implements a random variable generator that feeds from \( A_{HC} \) or \( A_{SZ} \) and provides as many samples as needed to generate synthetic \( \hat{A}_{HC} \) and \( \hat{A}_{SZ} \).

The random variable generator estimates the probability density functions (PDF) of each matrix column for healthy controls in \( f_i^c = \text{pdf}_{N}(\{A_{HC}_i\}) \) and schizophrenia patients in \( f_i^p = \text{pdf}_{N}(\{A_{SZ}_i\}) \), where \( i \) indicates the \( i^{th} \) column of the matrix, and the \( \text{pdf}_{N}(\cdot) \) function is a N-bin normalized histogram. Then, we follow a rejection sampling procedure for each PDF to form \( \hat{A}_{HC} \) and \( \hat{A}_{SZ} \). For example, in the case of generating healthy control images, we sample two uniformly distributed random variables \( u \sim U(0,1) \) and \( v \sim U(\min(\{A_{HC}_i\}), \max(\{A_{HC}_i\})) \) and keep \( v \) if \( u \leq f_i^c(v) \). This procedure is repeated until the required number of samples is accepted.
Table 1. Parameter settings to train the RBM at the pre-training stage and the feed-forward network at the discriminative fine-tuning stage.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Pre-training</th>
<th>Fine tuning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
<td>60465</td>
<td>60465</td>
</tr>
<tr>
<td>Unit type</td>
<td>Gaussian</td>
<td>Logistic</td>
</tr>
<tr>
<td>Dropout prob.</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>(l_1)-norm reg.</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.01</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Finally, we reconstruct \( M \) images for each group using the estimated source matrix from the original data and computing
\[
\hat{X}_{HC} = \hat{A}_{HC} [S] + \hat{X}, \text{ for healthy controls, and }
\hat{X}_{SZ} = \hat{A}_{SZ} [S] + \hat{X} \text{ for schizophrenia patients,}
\]
where, \( \hat{X} \) is the voxel mean computed at the beginning of the method. The method procedure is depicted in Fig. 1.

2.5. Experimental Setup

Our proposed approach uses synthetic data at the pre-training stage of the shallow network instead of using real data to do unsupervised learning (referred to hereafter as the regular approach). However, we evaluated both approaches to see if using synthetic data at pre-training can actually improve schizophrenia detection. The regular approach was applied to the same dataset used in this paper in [5]. In fact, we are recreating those results and comparing our method against the depth 1 model in [5]. The schizophrenia classification accuracy of these models was estimated using 10-fold cross-validation (CV) as follows.

GMC images were split into 10 approximately class-balanced folds. At each step, one fold was held out. The regular approach used the remaining 9 folds to train the RBM, whereas our proposed method used those folds to generate 5000 synthetic images (about 10 times the size of the original dataset) to be used at pre-training. After the unsupervised training stage, both approaches used the data from those 9 folds to fine-tune the network using backpropagation. This process was repeated 10 times, thus generating 10 models for each approach. The parameters used to train these models, which were retrieved from [5], are displayed in Table 1.

After models from both approaches were trained, they were tested using 10-fold CV and the same data partition scheme. At each round of CV, the hidden layer activations of 9 folds were used to train a set of classifiers, which were then evaluated on the left out (test) fold based on the F-score metric.

We used three conventional classification approaches: linear SVM, logistic regression (LR) and k-nearest neighbors (k-NN). [13]. These classifiers were tested using three data representations: raw data, the features extracted by the shallow network trained with the regular approach (pre-training with real data) and the ones estimated by the proposed approach (pre-training with extended synthetic dataset). We hypothesized that the performance of the classifiers trained with the regular approach features would be superior to the one achieved on raw data, while further improvement would be achieved using the proposed approach representation.

The parameters of the aforementioned classifiers were selected using a nested 5-fold CV procedure. The regularization parameters of SVM and LR were selected from a pool of 5 logarithmically spaced values in \([10^{-2}, 10^3]\), while for k-NN \(k \in \{1, 2, \ldots, 20\}\). The proposed shallow network was implemented using a GPU-based python implementation of deep learning algorithms (https://github.com/kiti/kiti/tree/master/convnet) and the classifiers were evaluated using the scikit-learn machine learning toolbox [14].

3. RESULTS

3.1. Classification Performance

Fig. 2 shows the boxplots of the F-scores achieved by SVM, LR and k-NN using raw data and the hidden layer data representations of the regular and the proposed approaches. By using the regular approach, the performance of the classifiers either remains the same as the one obtained using raw data (LR) or it improves slightly (SVM), achieving a considerable improvement for k-NN. On the other hand, the proposed model achieves a more consistent improvement over the classification performance on raw data and the features extracted by the regular approach. We tested the significance of the F-score improvement of the proposed approach over the regular one by means of a paired one-tailed \( t \)-test. Table 2 shows that the improvement was significant for all of the tested classifiers at 5% significance level.

3.2. Analysis of Receptive Fields

In order to have a better interpretation of which regions of the brain better characterized group spatial patterns of GMC and their differences between controls and patients according to our model, we analyzed the receptive field of the most discriminative hidden unit after pre-training and fine tuning, respectively. This unit had the highest \( t \)-score after performing a two-sample \( t \)-test on the hidden units activations for controls and patients on the entire dataset.

Fig. 3 (top) displays the hidden unit receptive field after pre-training. It can be seen that it spans the middle and inferior temporal
In addition to those regions, a small region in the right middle frontal gyrus, the fusiform gyrus, the superior and medial frontal gyri and the anterior cingulate cortex. The unit’s receptive field after fine tuning (bottom) shows much sparser weights, the nonzero ones being located on small clusters within the regions identified for pre-training. In addition, this work is an extension of [4] as it performs classification of MRI data on top of a pre-trained RBM. This paper also was obtained, providing evidence that a better characterization of groups can be achieved by using an extended dataset for pre-training. In addition, this work is an extension of [4] as it performs classification of MRI data on top of a pre-trained RBM. This paper also reveals the spatial extent of the receptive field of the most discriminative hidden unit for unsupervised and supervised learning. A remarkable finding is that several regions detected by our model (superior and medial frontal gyri; inferior temporal and fusiform gyri) are consistent with spatial components that characterize GMC deficits in schizophrenia reported in [15]. The cited publication analyzes patterns of GMC on healthy controls and schizophrenia patients in the largest aggregated structural image dataset to date.

While this work uses synthetic sMRI data for pre-training on a shallow network, this approach can be readily applied to deeper networks. Another possibility of future work that would be very useful for the neuroimaging field is to provide an intuitive interpretation of what kind of information is retrieved at each layer of deep architectures. By performing this analysis, complex brain networks would be revealed.

4. DISCUSSION AND CONCLUSIONS

This paper proposes a novel approach to generate synthetic sMRI data based on the ICA decomposition of real sMRI data to pre-train a shallow network model. By using this approach, an increased dataset can be used to estimate a better parameter initialization of the model for schizophrenia classification. This tackles the inherent problem of neuroimaging datasets being composed of a relatively small amount of data and makes them more suitable for deep learning applications.

Our methodogy improves the results obtained in [5] using the same schizophrenia dataset for shallow data representations. A reasonable average improvement on classification performance of 5% was obtained, providing evidence that a better characterization of groups can be achieved by using an extended dataset for pre-training. In addition, this work is an extension of [4] as it performs classification of MRI data on top of a pre-trained RBM. This paper also reveals the spatial extent of the receptive field of the most discriminative hidden unit for unsupervised and supervised learning. A remarkable finding is that several regions detected by our model (superior and medial frontal gyri; inferior temporal and fusiform gyri) are consistent with spatial components that characterize GMC deficits in schizophrenia reported in [15]. The cited publication analyzes patterns of GMC on healthy controls and schizophrenia patients in the largest aggregated structural image dataset to date.

Acknowledgement

This work was funded by NIH 5-R01-EB005846-10 grant and the Phase II COBRE project (NIH 5P20GM103472-07 grant).

5. REFERENCES