

Manifold Learning of Brain MRIs by Deep Learning

Tom Brosch^{1,3} and Roger Tam^{2,3}
for the Alzheimer's Disease Neuroimaging Initiative*

¹ Department of Electrical and Computer Engineering

² Department of Radiology

³ MS/MRI Research Group, University of British Columbia, Vancouver, Canada

Abstract. Manifold learning of medical images plays a potentially important role for modeling anatomical variability within a population with applications that include segmentation, registration, and prediction of clinical parameters. This paper describes a novel method for learning the manifold of 3D brain images that, unlike most existing manifold learning methods, does not require the manifold space to be locally linear, and does not require a predefined similarity measure or a prebuilt proximity graph. Our manifold learning method is based on deep learning, a machine learning approach that uses layered networks (called deep belief networks, or DBNs) and has received much attention recently in the computer vision field due to their success in object recognition tasks. DBNs have traditionally been too computationally expensive for application to 3D images due to the large number of trainable parameters. Our primary contributions are 1) a much more computationally efficient training method for DBNs that makes training on 3D medical images with a resolution of up to $128 \times 128 \times 128$ practical, and 2) the demonstration that DBNs can learn a low-dimensional manifold of brain volumes that detects modes of variations that correlate to demographic and disease parameters.

Keywords: Manifold Learning, Machine Learning, Brain Imaging, MRI, Deep Learning, Deep Belief Networks.

1 Introduction

The need for manifold learning often arises when very high-dimensional data needs to be analyzed but the intrinsic dimensionality of the data is much lower.

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI-Acknowledgement_List.pdf

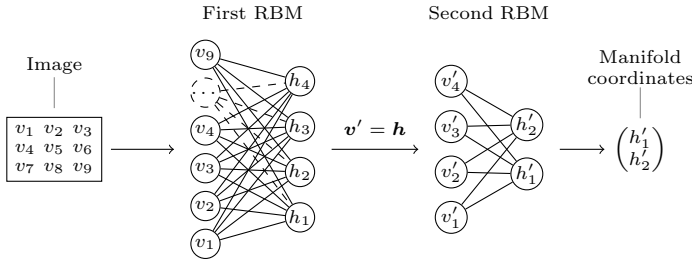


Fig. 1. Layer-wise dimensionality reduction network with 2 RBM layers. The visible units of the first layer are set to the image intensities. The hidden units of the last layer are the manifold coordinates.

This situation occurs, for example, when trying to visualize variability and common patterns in a given group of magnetic resonance images (MRIs) of the brain. Each image can be regarded as a point in a high-dimensional image space (called the ambient space), with $n_x \times n_y \times n_z$ coordinates, where n_x, n_y, n_z are the dimensions of each image. On the other hand, each image could also be identified by a smaller set of parameters that describe shape variations and patterns that are common for a particular group of images. These parameters span a new space called the manifold space. The task of manifold learning is to discover the low-dimensional space and its parameters which can then be used to model the anatomical variability within a population.

Various methods for manifold learning have been proposed (e.g., locally linear embedding (LLE) [1], Laplacian eigenmaps (LEM) [2], Isomaps [3]), with Isomaps and LEM being the most popular for medical image analysis. Both methods require a prebuild proximity graph. In order to build the proximity graph, it is assumed that the manifold space is locally linear, which means that distances between neighboring points in manifold space can be approximated by their distances in ambient space. Gerber et. al. have shown that the choice of a suitable distance measure is crucial for manifold learning using Isomaps and that the warping distance between brain images improves the learning performance over previously used Euclidean distances in the image space [4].

Manifolds have been used to regularize the segmentation of brain ventricles [5], and to constrain the deformable registration of brain images to have biologically plausible parameters [6]. Gerber et al. used Isomaps to predict clinical parameters of Alzheimer's (AD) patients [4], and Wolz et al. used Laplacian eigenmaps to perform biomarker discovery [7], also of AD patients, and atlas propagation for the segmentation of the hippocampus [8].

In this paper, we propose a novel approach for learning the manifold of brain images that uses a deep belief network (DBN) [9] to discover patterns of similarity in groups of images. In contrast to previous brain manifold learning algorithms, DBNs do not assume the manifold space to be locally linear and do not require a previously defined similarity measure or the construction of a proximity graph. Despite their popularity in computer vision (e.g., hand-written digit

classification [9], object recognition [10]), the use of DBNs has so far been largely limited to small 2D images due to the extremely high computational cost of training a DBN. The much higher resolution of medical images has traditionally made training of DBNs too slow to be practical. A recent advance in efficiency was the introduction of convolutional DBNs (convDBNs) [10], which has made training of 2D images with a resolution of up to 200×200 pixels feasible. Our primary contributions are 1) a much more computationally efficient training method for convDBNs that performs parameter learning in frequency domain in order to avoid the time consuming calculation of convolutions while minimizing the number of transformations to and from frequency space; when executed in parallel on graphical processors, training on images with a resolution of $128 \times 128 \times 128$ voxels can be done quite practically with low-cost consumer hardware, and 2) the demonstration that DBNs can learn a low-dimensional manifold of brain volumes that detects modes of variation that correlate to demographic and disease parameters.

2 Methods

Our proposed method performs manifold learning by reducing the dimensionality of the input images using a DBN, a deep generative model that is composed of multiple restricted Boltzmann machines (RBMs) [9] as illustrated by the simplified example in Fig. 1. An RBM is a Markov random field with trainable weights whose nodes are divided into a visible layer $\mathbf{v}$ representing the inputs of the model and a hidden layer $\mathbf{h}$ representing extracted features from the inputs. The first RBM receives the intensity values of a group of images as input and reduces the dimensionality of each image by discovering patterns of similarity that are common within groups of images. Subsequent RBMs receive the hidden unit activations of the previous RBM as input, thus learning successively more complex and abstract patterns from a training set. The number of trainable weights increases significantly with the resolution of the training images. In order to scale the model to high-resolution images, the first several layers of our DBN are convolutional RBMs (convRBMs), a type of RBM that uses weight sharing to reduce the number of trainable weights, albeit at the cost of using the much more computationally expensive convolutions instead of multiplications. In the remainder of this section, we will briefly review the training of convRBMs, followed by a description of our novel training algorithm that performs parameter learning in frequency domain. For comprehensive introductions to RBMs and convRBMs, the reader is referred to [9] and [10], respectively.

Our algorithm for the dimensionality reduction of an input image using a convRBM is illustrated in Fig. 2. In contrast to previous work that uses max pooling to reduce the dimensionality [10], all steps involved in our method are invertible, which allows the reconstruction of images from their manifold coordinates. In the first step, the pixels of an input image are reorganized into multiple images of lower resolution in order to reduce the dimensionality of a single image. The number of images in the image vector is then reduced with the following

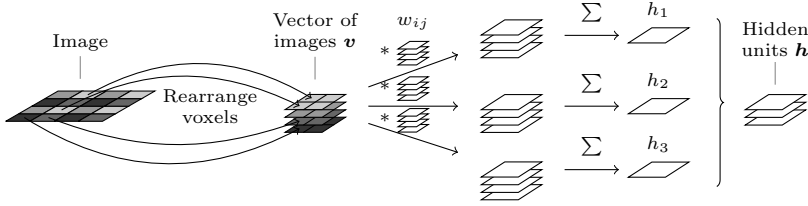


Fig. 2. Dimensionality reduction using a convRBM. We map neighboring voxels in an image to different images of lower dimensions, then apply convolutions and stack the sum of the images within each vector to form a new vector of images that represents the hidden units.

steps. To apply the model to real-valued data like the intensities of some medical images, the visible units are modeled as Gaussian units. When the visible units have been mean centered and standardized to unit variance, the expectation of the visible units is given by

$$\mathbb{E}[v_i | \mathbf{h}] = \sum_j w_{ij} * h_j + b_i \quad (1)$$

where $v_i, h_j, b_i \in \mathbb{R}^{N \times N \times N}$, b_i are bias terms, N is the image size, $w_{ij} \in \mathbb{R}^{N_w \times N_w \times N_w}$ are the weights, N_w is the size of a weight kernel, and $*$ denotes circular convolution. A binary hidden unit can only encode two states. In order to increase the expressive power of the hidden units, we use noisy rectified linear units as the hidden units, which has shown to improve the learning performance [11] of RBMs. The hidden units can be sampled with

$$h_j \sim \max(0, \mu_j + \mathcal{N}(0, \text{sigm}(\mu_j))) \quad (2)$$

where $\mu_j = \sum_i \tilde{w}_{ij} * v_i + c_j$, $c_j \in \mathbb{R}^{N \times N \times N}$, $\tilde{w}$ denotes a flipped version of w in all three dimensions, $\mathcal{N}(0, \sigma^2)$ denotes Gaussian noise, and $\text{sigm}(x)$ is the sigmoid function defined as $\text{sigm}(x) = (1 + \exp(-x))^{-1}$, $x \in \mathbb{R}$. The weights and bias terms of a convRBM can be learned using contrastive divergence (CD) [9]. During each iteration of the algorithm, the gradient of the parameters is estimated and a gradient step with a fixed learning rate is performed. The gradient of the weights can be approximated by:

$$\Delta w_{ij} = v_i * \tilde{h}_j - v'_i * \tilde{h}'_j \quad (3)$$

where h_j and h'_j are samples drawn from $p(h_j | \mathbf{v})$ and $p(h_j | \mathbf{v}')$ using (2) and $v'_i = \mathbb{E}[v_i | \mathbf{h}]$.

The computational bottleneck of the training algorithm is the calculation of convolutions, which has to be performed $4 \times V \times H$ times per iteration, where $V = |\mathbf{v}|$ and $H = |\mathbf{h}|$. To speed up the calculation of convolutions, we perform training in frequency domain, which maps the calculation of convolutions to simple element-wise multiplications. For example, this reduces the number of multiplications required to calculate the activation of one hidden unit with a weight

kernel size of 7 from $343 \times V$ to only V multiplications. In order to avoid the time-consuming calculation of Fourier transforms every time an image is convolved with a weight kernel, we map all operations needed for training to frequency domain wherever possible. Flipping of a convolutional kernel $\tilde{h}(a) = h(-a)$ can be expressed by element-wise calculation of the complex conjugate, which follows directly from the time-reversal property of the Fourier transform and the reality condition $\hat{f}(-\xi) = \overline{\hat{f}(\xi)}$. The only operations that can not be directly mapped to frequency domain are the calculation of the maximum function and the generation of Gaussian noise. To perform these operations, an image needs to be mapped to the spatial domain and back. However, these operations need only be calculated H times per iteration and are therefore not a significant contributor to the total runtime. Using the aforementioned mappings, equations (1) to (3) can be rewritten as

$$\mathbb{E}[\hat{v}_i | \hat{\mathbf{h}}] = \sum_j \hat{w}_{ij} \cdot \hat{h}_j + \hat{b}_i \quad (4)$$

$$\hat{h}_j \sim \mathcal{F}(\max(0, \mathcal{F}^{-1}(\hat{\mu}_j) + \mathcal{N}(0, \sigma^2))) \quad (5)$$

$$\Delta \hat{w}_{ij} = \hat{v}_i \cdot \overline{\hat{h}_j} - \hat{v}'_i \cdot \overline{\hat{h}'_j} \quad (6)$$

where $\hat{\mu}_j = \sum_i \overline{\hat{w}_{ij}} \cdot \hat{v}_i + \hat{c}_j$, $\sigma^2 = \text{sigm}(\mathcal{F}^{-1}(\hat{\mu}_j))$, $\hat{x} = \mathcal{F}(x)$ denotes x in frequency domain, $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, and $\cdot$ denotes element-wise multiplication. Training in frequency domain requires solely the calculation of element-wise operations and reductions, which can be efficiently performed on graphics cards.

3 Experiments and Results

We have evaluated the proposed method on a subset of the ADNI dataset [12], containing 300 T1-weighted MRIs of Alzheimer’s disease (AD) and normal subjects. The images were provided skull-stripped and bias field corrected. We re-sampled all images to a resolution of $128 \times 128 \times 128$ voxels and a voxel size of $2.0 \text{ mm} \times 2.0 \text{ mm} \times 2.0 \text{ mm}$. We then normalized their intensities to a common range, and rigidly registered them to a group-wise mean image prior to training and testing. We did not perform non-rigid registration for spatial normalization in order to evaluate the capabilities of the method without the added confound of complex registration parameters. The dataset was divided into a training set and a test set such that each set contains 75 AD and 75 normal subjects. To learn the manifold of brain MRIs, we used a DBN with three convRBM layers and two RBM layers. After three convRBMs, the dimension of each image is reduced to $8 \times 8 \times 8$ and small enough for RBMs. The training of the DBN took approximately 43 hours on two GeForce GTX 560 Ti graphics cards.

The geometric fit of the learned manifold model was evaluated in terms of the generalizability to new images and the specificity to images from the training set. The generalizability was measured as the average root mean squared error (RMSE) between an image and its reconstruction, normalized by the intensity

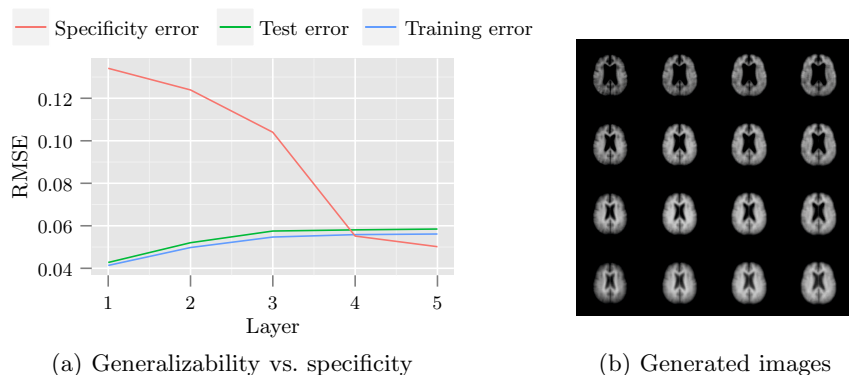


Fig. 3. (a) The similarity of the reconstruction errors between the training and test images indicates that no overfitting occurs. The opposite slopes of the reconstruction errors and error of generated images (specificity error) indicates a trade-off between generalizability vs. specificity in the earlier phases of training, but the low errors at Layer 5 indicate that the method is both generalizable and specific. (b) Axial slices from generated volumes from the manifold. An increase of the first and second manifold dimension visually correlates with an increase in brain and ventricle size, respectively.

range of the input image. The specificity was measured by calculating the average RMSE between images randomly generated from the manifold model and the most similar images from the training set. Figure 3(a) shows a comparison of the reconstruction errors between the training and test sets, and the specificity at different layers of the DBN. The similarity of the reconstruction errors between the training and test images indicates that no overfitting is occurring. The average reconstruction error at the last layer is below 6 %. Even though the very small reconstruction error is partially due to head MRIs having a large amount of homogeneous background, it demonstrates the ability of the learned manifold to capture most of the visual information with only two manifold parameters. The opposite slopes of the reconstruction errors and error of generated images indicates a trade-off between generalizability and specificity in the earlier phases of training. The low errors at the end of training (Layer 5) indicates that the method is able to be both specific and generalizable.

Figure 3(b) shows axial slices of 16 volumes sampled at the grid points of a 2D regular grid in manifold space. Volumes sampled along the first manifold dimension (from left to right) appear to increase in brain size, and the images sampled along the second manifold dimension (from bottom to top) appear to increase in ventricle size. Figure 4 shows an axial slice of each image of the training set plotted against its manifold coordinates. Consistent with images sampled from the manifold, an increase in ventricle size, which is indicative of brain atrophy (a hallmark of AD), visually correlates with an increase of the second manifold coordinate. The AD/normal status is indicated by the frame color of each image. The vertical separation between AD and normals suggests that the second manifold coordinate is potentially of practical use in differentiating between AD and normal.

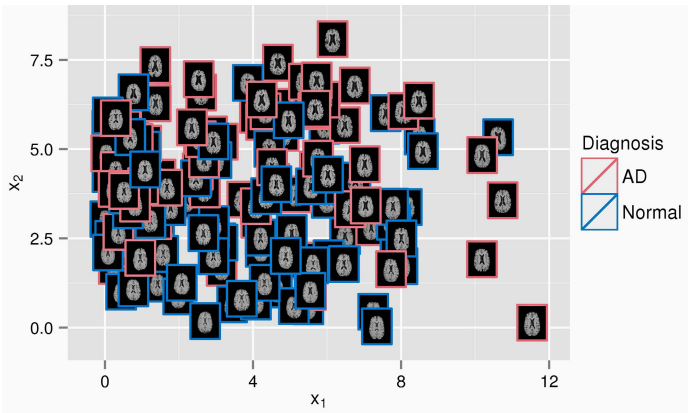


Fig. 4. Axial slices of volumes from the training set plotted against their manifold coordinates. The brains with larger ventricles, indicative of atrophy, are mostly at the top, which is also where most of the AD patients are.

Table 1. Pearson correlation r of demographic and clinical parameters with manifold coordinates (x_1, x_2) . The stronger correlation in each column is highlighted in bold.

	Age		Gender		MMSE		AD/Normal Status	
	r	p -value	r	p -value	r	p -value	r	p -value
x_1	-0.17	0.03	0.45	$8.24 \cdot 10^{-9}$	0.01	0.89	-0.03	0.69
x_2	0.45	$9.85 \cdot 10^{-9}$	0.19	0.02	-0.40	$3.53 \cdot 10^{-7}$	0.41	$1.54 \cdot 10^{-7}$

To evaluate the potential of the manifold coordinates to reveal or predict clinically relevant information, we have calculated the Pearson correlation r of demographic parameters (age, gender) and disease parameters (mini-mental state examination (MMSE) score, AD/normal status) with the manifold coordinates $(x_1$ and $x_2)$. The results of the correlation tests are summarized in Table 1. Age, MMSE and AD/normal status show highly significant correlations with x_2 (p -values between 9.85×10^{-9} and 3.53×10^{-7}), which makes intuitive sense because x_2 visually correlates with ventricle size. The first manifold coordinate correlates strongest with gender (p -value = 8.24×10^{-9}), which also makes sense in terms of the general difference in size between male and female. The strength and significance of the correlations demonstrate the potential of deep learning of brain images for classification and prediction of disease status.

4 Conclusions

We have proposed a novel approach to learning the manifold of brain MRIs. In contrast to previous work, our approach does not require an explicitly defined similarity measure, or building a proximity graph. Furthermore, we have shown that the learned manifold coordinates capture shape variations of the brain

that correlate with demographic and disease parameters. Our manifold learning method uses deep learning to discover patterns of similarity and variability within a group of images. Our proposed DBN training algorithm is much more efficient than traditional, convolution-based methods and, when parallelized on graphics processors, makes DBN training on 3D MRIs practical for the first time. For future work, we plan to incorporate clinical parameters into the training process to learn a model of the joint probability of images and associated clinical parameters. We would also like to investigate the effect of spatial normalization using affine or deformable registration, which we expect would result in further improvements in correlations with clinically relevant parameters. In addition, we would like to compare our method to other state-of-the-art brain manifold learning methods to investigate relative strengths and weaknesses for particular clinical applications such as the prediction of mild cognitive impairment to AD conversion, which is a topic of strong clinical interest.

Acknowledgements. This work was supported by internal funding from the University of British Columbia and grant funding from NSERC.

References

1. Saul, L., Roweis, S.: Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research* 4, 119–155 (2003)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6), 1373–1396 (2003)
3. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
4. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R.: Manifold modeling for brain population analysis. *Medical Image Analysis* 14(5), 643–653 (2010)
5. Etyngier, P., Ségonne, F., Keriven, R.: Shape priors using manifold learning techniques. In: 11th International Conference on Computer Vision, pp. 1–8. IEEE (2007)
6. Hamm, J., Ye, D., Verma, R.: Gram: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis* 14(5), 633–642 (2010)
7. Wolz, R., Aljabar, P., Hajnal, J.V., Lotjonen, J., Rueckert, D.: Manifold learning combining imaging with non-imaging information. In: IEEE International Symposium on Biomedical Imaging, pp. 1637–1640 (2011)
8. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: LEAP: learning embeddings for atlas propagation. *NeuroImage* 49(2), 1316–1325 (2010)
9. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. *Neural Computation* 18(7), 1527–1554 (2006)
10. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM* 54(10), 95–103 (2011)
11. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th Annual International Conference on Machine Learning, pp. 807–814 (2010)
12. Petersen, R., Aisen, P., Beckett, L., Donohue, M., Gamst, A., et al.: Alzheimer’s disease neuroimaging initiative (ADNI). *Neurology* 74(3), 201–209 (2010)