Disentangled Sequential Graph Autoencoder for Preclinical Alzheimer's Disease Characterizations from ADNI Study

Fan Yang[1(✉)], Rui Meng[2], Hyuna Cho[3], Guorong Wu[4], and Won Hwa Kim[1,3(✉)]

[1] University of Texas at Arlington, Arlington, USA
[2] Lawrence Berkeley National Laboratory, Berkeley, USA
[3] Pohang University of Science and Technology, Pohang, South Korea
wonhwa@postech.ac.kr
[4] University of North Carolina, Chapel Hill, Chapel Hill, USA

Abstract. Given a population longitudinal neuroimaging measurements defined on a brain network, exploiting temporal dependencies within the sequence of data and corresponding latent variables defined on the graph (i.e., network encoding relationships between regions of interest (ROI)) can highly benefit characterizing the brain. Here, it is important to distinguish time-variant (e.g., longitudinal measures) and time-invariant (e.g., gender) components to analyze them individually. For this, we propose an innovative and ground-breaking Disentangled Sequential Graph Autoencoder which leverages the Sequential Variational Autoencoder (SVAE), graph convolution and semi-supervising framework together to learn a latent space composed of time-variant and time-invariant latent variables to characterize disentangled representation of the measurements over the entire ROIs. Incorporating target information in the decoder with a supervised loss let us achieve more effective representation learning towards improved classification. We validate our proposed method on the longitudinal cortical thickness data from Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Our method outperforms baselines with traditional techniques demonstrating benefits for effective longitudinal data representation for predicting labels and longitudinal data generation.

1 Introduction

Representation learning is at the core of Image Analysis. Lots of recent attentions are at a disentangled representation of data, as the individual disentangled representations are highly sensitive to a specific factor whereas indifferent to others [2,10,13,23,30]. A typical disentangling method would find a low-dimensional

F. Yang and R. Meng are joint first authors.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12902, pp. 362–372, 2021.
https://doi.org/10.1007/978-3-030-87196-3_34

latent space for high-dimensional data whose individual latent dimensions correspond to independent disentangling factors. For longitudinal data, one can expect to decompose the longitudinal data into time-invariant factors and time-variant factors by obtaining the "disentangled" representation as longitudinal observations are affected by both time-variant and static variables [12,19,31]. In the context of neuroimaging studies, the disentangled representation would be able to separate time-independent concepts (e.g. anatomical information) from dynamical information (e.g. modality information) [25], which may offer effective ways of compression, conditional data generation, classification and others.

Recent advances in variational autoencoders (VAE) [16] have made it possible to learn various representations in an unsupervised manner for neuroimaging analysis [1,30]. Moreover, various vibrant of autoencoders are also proposed to model temporal data; for example, [12] introduced the factorised hierarchical variational auto-encoder (FHVAE) for unsupervised learning of disentangled representation of time series. Sequential variational autoencoder was proposed in [19] benefiting from the usage of the hierarchical prior. It disentangles latent factors by factorizing them into time-invariant and time-dependent parts and applies an LSTM sequential prior to keep a sequential consistency for sequence generation. [31] modeled the time-varying variables via LSTM in both encoder and decoder for dynamic consistency.

There are two major issues with current approaches. First, while these methods can deal with temporal nature of the data, they do not necessarily introduce supervision at all. Moreover, from a neuroscience perspective, the domain knowledge tells us that the regions of interest (ROIs) in the brain are highly associated to each other both functionally and structurally [7,17,18,20]. This association provides a prior knowledge on connection between the ROIs as a graph; for example, structural brain connectivity from tractography on Diffusion Tensor Imaging (DTI) provides a path for anisotropic variation and diffusion of structural changes in the brain such as atrophy of cortical thickness. Most of the existing methods do not consider this arbitrary topology of variables, if there is any, into account, which can provide significant benefit for downstream tasks. To summarize, learning with (either full or partial) supervision on longitudinal neuroimaging measurements on a brain network is still **under-explored**.

Given longitudinal observations (e.g., cortical thickness) on specific ROIs in the brain and a structural brain network characterized by bundles of neuron fiber tracts, our aim is to develop a framework to learn a latent disentangled representation of the observations that are composed of time-variant and time-invariant latent variables. For this, we propose an innovative Semi-supervised Sequential Graph Autoencoder model which leverages ideas from the sequential variational autoencoder (SVAE), graph convolution and semi-supervising framework. The core idea is to incorporate target information as a supervision in the decoder with a supervised loss, which let us achieve more effective representation for downstream tasks by balancing extraction of underlying structure as well as accurately predicting class labels.

Our proposed framework learns a latent disentangled representation composed of time-variant and time-invariant latent variables to characterize the

longitudinal measurements over the entire structural brain network that consists of ROIs. Our **contributions** are as summarized follows: our model can 1) learn an ideal disentangled representation which separates time-independent content or anatomical information from dynamical or modality information and conditionally generate synthetic sequential data; 2) perform semi-supervised tasks which can jointly incorporate supervised and unsupervised data for classification tasks; 3) leverage graph structure to robustly learn the disentangling latent structure. Using our framework, we analyzed longitudinal cortical thickness measures on brain networks with diagnostic labels of Alzheimer's Disease (AD) from Alzheimer's Disease Neuroimaging Initiative (ADNI) study. As AD is a progressive neurodegenerative condition characterized by neurodegeneration in the brain caused by synthetic factors [6,14,22,27,28], it is important to effectively characterize early symptoms of the disease. We expect that disentangling ROI measures with time-variant and static components can provide unique insights.

2 Background

Our proposed framework involves two important concepts: 1) graph convolutions and 2) SVAE. Hence, we begin with brief reviews of their basics.

Graph Convolutions. Let $G = \{\mathbb{V}, \mathbb{E}, A\}$ be an undirected graph, where $\mathbb{V}$ is a set of nodes with $|\mathbb{V}| = n$, $\mathbb{E}$ is a set of edges and A is an adjacent matrix that specify connections between the nodes. Graph Fourier analysis relies on the spectral decomposition of graph Laplacian defined as $\mathcal{L} = D - A$, where D is a diagonal degree matrix with $D_{i,i} = \sum_j A_{i,j}$. The normalized Laplacian is defined as $L = I_n - D^{-1/2}AD^{-1/2}$, where I_n is the identity matrix. Since L is real and positive semi-definite, it has a complete set of orthonormal eigenvectors $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ with corresponding non-negative real eigenvalues $\{\lambda_l\}_{l=1}^n$. Eigenvectors associated with smaller eigenvalues carry slow varying signals, indicating that connected nodes share similar values. In contrast, eigenvectors associated with larger values carry faster varying signals across the connected nodes. We are interested in the smallest eigenvalues due to the negation used to compute the Laplacian matrix in terms of the Euclidean Commute Time Distance [26]. Let $\boldsymbol{x} \in \mathbb{R}^n$ be a signal defined on the vertices of the graph. The graph Fourier transform of $\boldsymbol{x}$ is defined as $\hat{\boldsymbol{x}} = U^T\boldsymbol{x}$, with inverse operation given by $\boldsymbol{x} = U\hat{\boldsymbol{x}}$. The graphical convolution operation between signal $\boldsymbol{x}$ and filter $\boldsymbol{g}$ is

$$\boldsymbol{g} * \boldsymbol{x} = U((U^T\boldsymbol{g}) \odot (U^T\boldsymbol{x})) = U\hat{G}U^T\boldsymbol{x}. \tag{1}$$

Here, $U^T\boldsymbol{g}$ is replaced by a filter $\hat{G} = \mathrm{diag}(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta} \in \theta^n$ in Fourier domain. Unfortunately, eigendecomposition of $\boldsymbol{L}$ and matrix multiplication with U are expensive. Motivated by the Chebyshev polynomials approximation in [7,9] introduced a Chebyshev polynomial paramterization for ChebyNet that offers fast localized spectral filtering. Later, [17] provided a simplified version of ChebyNet by considering a second order approximation such that $\boldsymbol{g} * \boldsymbol{x} \approx \theta(I_n + D^{-1/2}AD^{-1/2})\boldsymbol{x}$ and illustrate promising model performance in graph-based semi-supervising learning tasks, and GCN is deeply studied in [18]. Then,

FastGCN was proposed in [4] which approximates the original convolution layer by Monte Carlo sampling, and recently, [29] leveraged graph wavelet transform to address the shortcomings of spectral graph convolutional neural networks.

Sequential Variational Autoencoder. Variational autoencoder (VAE), initially introduced in [16] as a class of deep generative mode, employs a reparameterized gradient estimator for a evidence lower bound (ELBO) while applying amortized variational inference to an autoencoder. It simultaneously trains both a probabilistic encoder and decoder for elements of a data set $\mathcal{D} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ with latent variable $\boldsymbol{z}$. Sequential variational autoencoders (SVAEs) extend VAE to sequential data $\mathcal{D}$, where each data are $\boldsymbol{x}_{1:T} = (x_1, \ldots, x_T)$ [19,31]. SVAEs factorize latent variables into two disentangled variables: the time-invariant variable $\boldsymbol{f}$ and time-varying variable $\boldsymbol{z}_{1:T} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T)$. Accordingly, decoder is casted as a conditional probabilistic density $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{f}, \boldsymbol{z}_{1:T})$ and encoder is used to approximate the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{f}, \boldsymbol{z}_{1:T}|\boldsymbol{x})$ as $q_{\boldsymbol{\phi}}(\boldsymbol{f}, \boldsymbol{z}_{1:T}|\boldsymbol{x})$ that is referred to as an "inference network" or a "recognition network". $\boldsymbol{\theta}$ refer to the model parameters of generator and $\boldsymbol{\phi}$ refer to the model parameters of encoder. SVAEs are trained to maximize the following ELBO:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{D}) = \mathbb{E}_{\hat{p}(\boldsymbol{x}_{1:T})}\Big[\mathbb{E}_{q_\phi(\boldsymbol{z}_{1:T}, \boldsymbol{f}|\boldsymbol{x}_{1:T})} \ln p_\theta(\boldsymbol{x}_{1:T}|\boldsymbol{f}, \boldsymbol{z}_{1:T}) - \mathrm{KL}(q_\phi(\boldsymbol{f}, \boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T}), p_\theta(\boldsymbol{f}, \boldsymbol{z}_{1:T}))\Big], \tag{2}$$

where $\hat{p}(\boldsymbol{x}_{1:T})$ is the empirical distribution with respect to the data set $\mathcal{D}$, $q_{\boldsymbol{\phi}}(\boldsymbol{f}, \boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})$ is the variational posterior, $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}|\boldsymbol{f}, \boldsymbol{z}_{1:T})$ is the conditional likelihood and $p_{\boldsymbol{\theta}}(\boldsymbol{f}, \boldsymbol{z}_{1:T})$ is prior over the latent variables.

3 Proposed Model

Let us first formalize the problem setting. Consider a dataset consists of shared graph G, and M unsupervised data points $\mathcal{D} = \{\boldsymbol{X}_i\}_{i=1}^{M}$ and M^{sup} supervised data points $\mathcal{D}^{sup} = \{\boldsymbol{X}_i, y_i\}_{i=1}^{M^{sup}}$ as pairs. $\boldsymbol{X}_i = (X_{i,1}, \ldots X_{i,T_i})$ refer to the i-th sequential observations on N nodes of a graph G with C input channels, i.e., $X_{i,t} \in \mathbb{R}^{N \times C}$, and y_i is the corresponding class label such as diagnostic labels.

We propose a semi-supervised sequential variational autoencoder model, and for convenience we omit the index i whenever it is clear that we are referring to terms associated with a single data point and treat individual data as $(\boldsymbol{X}, y)$.

Objective Function. Typical semi-supervised learning pipelines for deep generative models, e.g., [15,24], define an objective function for optimization as

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{D}, \mathcal{D}^{sup}) = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{D}) + \tau \mathcal{L}^{sup}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{D}^{sup}). \tag{3}$$

Similarly, our approach jointly models unsupervised and supervised collections of terms over $\mathcal{D}$ and $\mathcal{D}^{sup}$. The formulation in Eq. 3 introduces a constant τ to control the relative strength of the supervised term. As the unsupervised term in Eq. 3 is exactly same as that of Eq. 2, we focus on the supervised term $\mathcal{L}^{sup}$ in Eq. 3 expanded below. Incorporating a weighted component as in [15],

$$\mathcal{L}^{sup}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{D}^{sup}) = \mathbb{E}_{\hat{p}(\boldsymbol{X}, y)}\left[\mathbb{E}_{q_\phi(\boldsymbol{f}, \boldsymbol{z}|\boldsymbol{X}, y)}\left[\ln \frac{p_\theta(\boldsymbol{X}, y, \boldsymbol{f}, \boldsymbol{z})}{q_\phi(\boldsymbol{f}, \boldsymbol{z}|\boldsymbol{X}, y)}\right] + \alpha \ln q_\phi(y|\boldsymbol{X})\right] \tag{4}$$

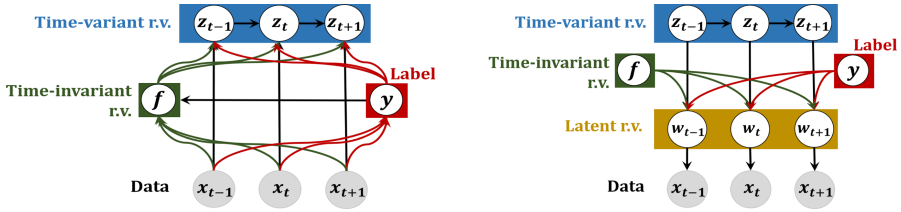

Fig. 1. A graphical model visualisation of the encoder (left) and decoder (right). In the encoder, label y is inferred by data $\boldsymbol{x}$ and time-invariant r.v. $\boldsymbol{f}$ are inferred by label y and data $\boldsymbol{x}$, and time-varying r.v. $\boldsymbol{z}$ are sequentially inferred by label y, time-invariant r.v. $\boldsymbol{f}$ and data $\boldsymbol{x}$. In the decoder, data are sequentially generated from time-invariant random variable (r.v.) $\boldsymbol{f}$, time-varying r.v. $\boldsymbol{z}$ and label y via latent r.v. $\boldsymbol{w}$.

where α balances the classification performance and reconstruction performance. Discussions on generative and inference model will continue in the later sections.

Generative Model. This section discusses modeling conditional probabilistic density $p_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{f}, \boldsymbol{z}, y)$ with its corresponding prior. We incorporate the topology information of the graph G into the generative process using a graph convolution. Specifically, we assume that sequences $\boldsymbol{X}$ are generated from P-dimensional latent vectors $\boldsymbol{W} = (W_1, \ldots, W_T)$ and $W_t \in \mathbb{R}^{N \times P}$ via

$$X_t = \hat{A} W_t \Theta, \tag{5}$$

where $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$, $\tilde{A} = A + I$ and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. A is the adjacent matrix for the graph G and Θ is the trainable weight matrix. Then we assume the latent variables $\boldsymbol{W}$ are generated from two disentangled variables: the time-invariant (or static) variable $\boldsymbol{f}$ and the time-varying (or dynamic) variables $\boldsymbol{z}$, as well as label $\boldsymbol{y}$, as shown in Fig. 1. A joint for the generative model is given as

$$p_{\boldsymbol{\theta}}(\boldsymbol{X}, y, \boldsymbol{z}, \boldsymbol{f}) = p_{\boldsymbol{\theta}}(\boldsymbol{f}) p_{\boldsymbol{\theta}}(y) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{z}_t|\boldsymbol{z}_{<t}) p_{\boldsymbol{\theta}}(X_t|y, \boldsymbol{f}, \boldsymbol{z}_t). \tag{6}$$

The prior of $\boldsymbol{f}$ is defined as a Gaussian distribution: $\boldsymbol{f} \sim \mathcal{N}(\mathbf{0}, I)$. Time-varying latent variables $\boldsymbol{z}_{1:T}$ follow a sequential prior $\boldsymbol{z}_t|\boldsymbol{z}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_t, \mathrm{diag}(\boldsymbol{\sigma}_t^2))$, where $[\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t]$ are estimated by a recurrent network, such as LSTM [11] or GRU [5], in which the hidden states are updated temporally. The generating distribution of W_t is conditional on y, $\boldsymbol{f}$ and $\boldsymbol{z}_t$: $\mathrm{vec}(W_t)|y, \boldsymbol{f}, \boldsymbol{z}_t \sim \mathcal{N}(\boldsymbol{\mu}_{w,t}, \mathrm{diag}(\boldsymbol{\sigma}_{w,t}^2))$, where $[\boldsymbol{\mu}_{w,t}, \boldsymbol{\sigma}_{w,t}] = \psi^{Decoder}(y, \boldsymbol{f}, \boldsymbol{z}_t)$. This decoder $\psi^{Decoder}$ can be any flexible neural network such as multilayer perceptron (MLP). The $\boldsymbol{f}$ will be capable of modelling global aspects of the whole sequences which are time-invariant, while $\boldsymbol{z}_{1:T}$ will model time-varying features. As mentioned in [19], to separate the static and dynamic information, smaller dimension of $\boldsymbol{z}_t$ is preferred. In the context of ADNI study, $\boldsymbol{z}_t$ would encode how ROIs at timestamp t is morphed into those at timestamp $t+1$. In the context of generative model, we employ LSTM as the

prior for $\boldsymbol{z}$ and use MLP for the conditional probabilistic density, and we set the dimension $P = 1$.

Inference Model. The developed SVAE within our framework proposes a recognition model $q_{\boldsymbol{\phi}}(y, \boldsymbol{f}, \boldsymbol{z}|\boldsymbol{X}) = q_{\boldsymbol{\phi}}(y|\boldsymbol{X})q_{\boldsymbol{\phi}}(\boldsymbol{f}, \boldsymbol{z}|y, \boldsymbol{X})$ to approximate the posterior $p_{\boldsymbol{\theta}}(y, \boldsymbol{f}, \boldsymbol{z}|\boldsymbol{X})$. The recognition model is formulated as

$$y \sim \mathrm{Cat}(\mathrm{Softmax}(\boldsymbol{p}_y)), \quad \boldsymbol{f} \sim \mathcal{N}(\boldsymbol{\mu}_f, \mathrm{diag}(\boldsymbol{\sigma}_f^2)), \quad \boldsymbol{z}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \mathrm{diag}(\boldsymbol{\sigma}_t^2)), \tag{7}$$

where $\boldsymbol{p}_y = \psi_y^{Encoder}(X_{1:T})$, $[\boldsymbol{\mu}_f, \boldsymbol{\sigma}_f] = \psi_f^{Encoder}(y, X_{1:T})$ and $[\boldsymbol{u}_t, 2\log\boldsymbol{\sigma}_t] = \psi_R^{Encoder}(y, X_{\leq t})$. It implies that the label y and the time-invariant variable $\boldsymbol{f}$ are conditional on the whole sequence via $\psi_y^{Encoder}$ and $\psi_f^{Encoder}$, while the time-dependent variable $\boldsymbol{z}_t$ is inferred by the sequence before time t, $X_{\leq t}$. The inference model is visualized in Fig. 1 and is factorized as

$$q_{\boldsymbol{\phi}}(y, \boldsymbol{z}_{1:T}, \boldsymbol{f}|X_{1:T}) = q_{\boldsymbol{\phi}}(y|X_{1:T})q_{\boldsymbol{\phi}}(\boldsymbol{f}|y, X_{1:T})\prod_{t=1}^{T} q_{\boldsymbol{\phi}}(\boldsymbol{z}_t|y, X_{\leq t}). \tag{8}$$

In the context of our inference model, we employ three independent LSTMs for three conditional probabilistic densities of y, $\boldsymbol{f}$ and $\boldsymbol{z}$.

4 Experimental Results

We conducted experiments on structural brain connectivity from DTI in ADNI. DTI images were processed by tractography, which extracted neuron fiber tracts and longitudinal cortical thickness measures registered at Destrieux atlas [8] with 148 ROIs. The dataset had five labels; we merged control (CN), Significant Memory Concern (SMC) and Early Mild Cognitive Impairment (EMCI) groups as Pre-clinical AD group, and Late Mild Cognitive Impairment (LMCI) and Alzheimer's Disease (AD) as Prodromal AD group to ensure sufficient sample

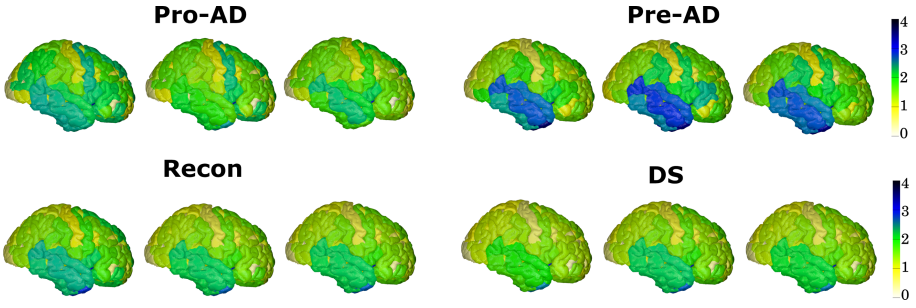

Fig. 2. Top panel shows the true brain surfaces at timestamp t_0, t_1 and t_2 for subject 1 (Pro-AD) and subject 2 (Pre-AD), respectively. Bottom panel shows the reconstructed brain surfaces for subject 1 (Recon) and subject 1's brain surfaces through the dynamic swapping (DS). Drawings generated using BrainPainter [21].

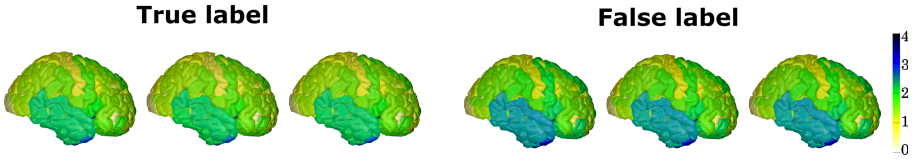

Fig. 3. Label swapping task. Left panel shows generated brain surfaces for subject 1 (Pro-AD) based on the true label at timestamp t_0, t_1 and t_2, respectively. Right panel shows generated brain surfaces for the same subject 1 but based on the false label.

size. The dataset included N = 140 subjects with the Pre-AD group (93 subjects/330 records) and the Pro-AD group (47 subjects/170 records). The mean (std) of ages and sex ratio (Male:Famale) in Pre-AD group and Pro-AD group are 74.02(6.72)/(185:145) and 74.87(6.92)/(95:75), respectively. An overall graph was obtained by taking the average of the adjacency matrices. Experiments for disentangle representation and quantitative analysis were performed given below.

4.1 Disentangled Representation

In this experiment, we randomly took 100 subjects' records for training, 20 subjects' records for validation and the other 20 subjects' records for testing. We set the dimension size of $\boldsymbol{f}$ as 8 and the dimension size of $\boldsymbol{z}$ as 32. We also set the size of hidden states in LSTMs as 32.

We randomly selected two subjects with more than three records (i.e., timepoints), where subject 1 belongs to Prodromal AD group and subject 2 belongs to Pre-clinical AD group. Suppose that the two subjects' sequential records are given for anatomical information and modality information denoted by R_1 and R_2. Our method performs the reconstruction task and the dynamic swapping task in which the record generation is based on the true y, $\boldsymbol{f}$ from R_1 and $\boldsymbol{z}$ from R_2 as in Fig. 2. It shows that the reconstruction captures both anatomical information and modality information, and figures generated from the dynamic swapping task illustrate that time-varying latent variables $\boldsymbol{z}$ succeed to learning the modality information.

In Fig. 3, we show results from the label swapping task on subject 1, where we generate cortical thickness based on the $\boldsymbol{f}$ from R_1, $\boldsymbol{z}$ from R_1 and true/false labels y. Comparing the generated measures of subject 1 with the true measures in Fig. 2, we found that generated measures based on the true label are more similar to the true measures and that based on the false label has totally different patterns but similar to the true measures of subject 2 in Fig. 2. It suggests that the decoder in our model correctly learns the label.

To understand the disentangled representation on the time invariant latent variable $\boldsymbol{f}$, we carry out latent traversals in $\boldsymbol{f}$ as in [3]. Specifically, we first computed the average Kullback–Leibler divergence for $\boldsymbol{f}$ with its prior. Then we selected the two dimensions in $\mathbf{f}$ with the largest two values (the 1st and 3rd elements), which refer to the two most informative dimensions and then

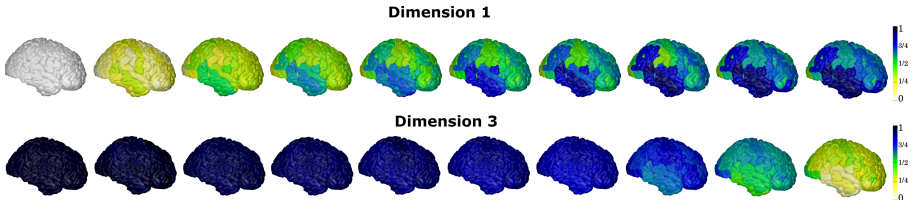

Fig. 4. Latent traversals task. Top: the latent brain surfaces for dim-1 on subject 1 (pro-AD). Bottom: the latent brain surfaces for dim-3 on subject 1.

traverse a single latent dimension on 10 equally spaced grids on $[-3, 3]$. For better visualization, we chose the first image as baseline and subtracted the baseline of image from all generated images. Then we normalized those images in a unit region $[0, 1]$ shown in Fig. 4.

4.2 Quantitative Analysis

We carry out 7-fold cross validation (CV) in which we take six folds for training (one fold for validation from the training set) and one fold for testing. We set the dimension size of $\boldsymbol{f}$ as 8 and the dimension size of $\boldsymbol{z}$ as 4. We also set the size of hidden states in LSTMs as 8. We compared our model with S3VAE model [31], which has a generator as in Fig. 1 but without a probabilistic model on label y. As S3VAE is unsupervised, we cannot directly compare our model with it. Instead, we tackle the classification task via a two-stage approach. Specifically, we train S3VAE to obtain latent $\boldsymbol{f}$ and train a naive neural network for the label classification. As for testing, we first get $\boldsymbol{f}$ from trained S3VAE and classify $\boldsymbol{f}$. Also, we propose a supervised loss based on the latent time-invariant f for S3VAE as one competitor. The generative model is modified as

$$p_{\theta}(\boldsymbol{X}, y, \boldsymbol{z}, \boldsymbol{f}) = p_{\theta}(\boldsymbol{f}) p_{\theta}(y|\boldsymbol{f}) \prod_{t=1}^{T} p_{\theta}(\boldsymbol{z}_t|\boldsymbol{z}_{<t}) p_{\theta}(X_t|\boldsymbol{f}, \boldsymbol{z}_t), \tag{9}$$

where we employ a fully connected network following a softmax activation function for $p_{\boldsymbol{\theta}}(y|\boldsymbol{f})$. We treat the pro-AD as positive result and then report three classification measures, accuracy, precision and recall. We also report root mean

Table 1. Mean (Std) reconstruction and classification performance with 7-fold cross validation.

	RMSE	Accuracy	Precision	Recall
Our model ($\alpha = 1$)	0.257 (0.041)	0.657 (0.168)	0.416 (0.367)	0.446 (0.349)
Our model ($\alpha = 10$)	0.258 (0.046)	0.736 (0.151)	0.541 (0.346)	0.492 (0.337)
S3VAE (Supervised)	0.263 (0.042)	0.664 (0.164)	0.000 (0.000)	0.000 (0.000)
S3VAE (Two stages)	0.254 (0.043)	0.664 (0.164)	0.000 (0.000)	0.000 (0.000)

square error (RMSE) as a reconstruction measure for testing data in Table 1. As for our proposed model, we consider the regularization weights $\alpha = 1$ and $\alpha = 10$. We find that our model has a better reconstruction performance in comparison to the supervised S3VAE model and performs similarly to the two-stage S3VAE. As for classification, our model with $\alpha = 10$ outperforms other models. We note that S3VAE based methods always categorize patients into pre-AD group, suffering from the imbalance classification issue. Our model resolves this issue and obtains a significantly better classification result according to both higher precision and recall scores. Finally, we note that to get better reconstruction or prediction results, properly tuning the hyperparameter α is important.

5 Conclusion

In summary, we propose a novel Sequential Autoencoder model. It incorporates the graph information via graph convolution operation, and it jointly models supervised and unsupervised data. Our model is flexible for data generation and it can conditionally generate sequential data based on label, disentangled time-invariant and time-varying latent variables. Quantitatively, we show that this model has competitive classification and reconstruction performance compared with two modified state-of-the-art S3VAE models.

Acknowledgement. This work was supported by GAANN Doctoral Fellowships in Computer Science and Engineering at UTA sponsored by the U.S. Department of Education, NSF IIS CRII 1948510, NIH RF1 AG059312, NIH R03 AG070701, and IITP-2019-0-01906 funded by MSIT (AI Graduate School Program at POSTECH).

References

1. Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In: Crimi, A., et al. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 161–169. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_16
2. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
3. Burgess, C.P., et al.: Understanding disentangling in beta-VAE. arXiv preprint arXiv:1804.03599 (2018)
4. Chen, J., Ma, T., Xiao, C.: Fastgcn: fast learning with graph convolutional networks via importance sampling. arXiv preprint arXiv:1801.10247 (2018)
5. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP, pp. 1724–1734. ACL (2014)
6. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y.: ADNI: individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage **59**(3), 2217–2230 (2012)
7. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Adv. Neural Inf. Process. Syst. **29**, 3844–3852 (2016)

8. Destrieux, C., Fischl, B., Dale, A., Halgren, E.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage **53**(1), 1–15 (2010)
9. Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. Appl. Comput. Harm. Anal. **30**(2), 129–150 (2011)
10. Higgins, I., et al.: beta-VAE: Learning basic visual concepts with a constrained variational framework (2016)
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
12. Hsu, W.N., Zhang, Y., Glass, J.: Unsupervised learning of disentangled and interpretable representations from sequential data. In: NeurIPS, pp. 1878–1889 (2017)
13. Kim, H., Mnih, A.: Disentangling by factorising. In: International Conference on Machine Learning, pp. 2649–2658. PMLR (2018)
14. Kim, W.H., Racine, A.M., Adluru, N., et al.: Cerebrospinal fluid biomarkers of neurofibrillary tangles and synaptic dysfunction are associated with longitudinal decline in white matter connectivity: a multi-resolution graph analysis. NeuroImage Clin. **21**, 101586 (2019)
15. Kingma, D.P., Mohamed, S., Rezende, D.J., Welling, M.: Semi-supervised learning with deep generative models. In: Advances in Neural Information Processing Syst. pp. 3581–3589 (2014)
16. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. CoRR abs/1312.6114 (2014)
17. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings. OpenReview.net (2017)
18. Li, Q., Han, Z., Wu, X.M.: Deeper insights into graph convolutional networks for semi-supervised learning. In: AAAI, vol. 32 (2018)
19. Li, Y., Mandt, S.: Disentangled sequential autoencoder. arXiv preprint arXiv:1803.02991 (2018)
20. Ma, X., Wu, G., Hwang, S.J., Kim, W.H.: Learning multi-resolution graph edge embedding for discovering brain network dysfunction in neurological disorders. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 253–266. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_20
21. Marinescu, R.V., Eshaghi, A., Alexander, D.C., Golland, P.: Brainpainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes. In: Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy, pp. 112–120. Springer (2019), https://doi.org/10.1007/978-3-030-33226-6
22. McKhann, G.M., et al.: The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Sementia **7**(3), 263–269 (2011)
23. Meng, R., Bouchard, K.: Bayesian inference in high-dimensional time-series with the orthogonal stochastic linear mixing model. arXiv preprint arXiv:2106.13379 (2021)
24. Siddharth, N., et al.: Learning disentangled representations with semi-supervised deep generative models. In: Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. (2017)

25. Ouyang, J., Adeli, E., Pohl, K.M., Zhao, Q., Zaharchuk, G.: Representation Disentanglement for Multi-modal MR Analysis. arXiv e-prints arXiv:2102.11456 (Feb 2021)
26. Saerens, M., Fouss, F., Yen, L., Dupont, P.: The principal components analysis of a graph, and its relationships to spectral clustering. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 371–383. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30115-8_35
27. Thompson, P.M., Hayashi, K.M., Sowell, E.R., Gogtay, N., Giedd, J.N., Rapoport, J.L., De Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., et al.: Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. Neuroimage **23**, S2–S18 (2004)
28. Wolz, R., et al.: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS ONE **6**(10), e25446 (2011)
29. Xu, B., Shen, H., Cao, Q., Qiu, Y., Cheng, X.: Graph wavelet neural network. In: International Conference on Learning Representations (2019)
30. Zhao, Q., Adeli, E., Honnorat, N., Leng, T., Pohl, K.M.: Variational autoencoder for regression: application to brain aging analysis. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 823–831. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_91
31. Zhu, Y., Min, M.R., Kadav, A., Graf, H.P.: S3VAE: self-supervised sequential VAE for representation disentanglement and data generation. In: CVPR, pp. 6538–6547 (2020)