(TMI 2021)
Dual Attention Multi-Instance Deep Learning for Alzheimer’s Disease Diagnosis With Structural MRI

Wenyong Zhu , Liang Sun , Jiashuang Huang , Liangxiu Han , and Daoqiang Zhang , Member, IEEE

Luoyao Kang
2021.12.15

Introduction

Structural magnetic resonance imaging (sMRI) is widely used for the brain neurological disease diagnosis, which could reflect the variations of brain. However, due to the local brain atrophy**,** only a few regions in sMRI scans have obvious structural changes, which are highly correlative with pathological features. Hence, the key challenge of sMRI-based brain disease diagnosis is to enhance the identification of discriminative features.

Motivation

Since brain atrophy usually occurs locally, only a few regions in sMRI scans have obvious structural changes which are highly correlative with pathological features, while the rest of regions have little useful information for distinction. Therefore, the key challenge of deep learning-based diagnosis with sMRI is to enhance the identification of discriminative features, including

1) informative micro-structures within local regions and

2) relatively important regions in a global image.

Contributions

1) A dual attention multi-instance deep learning model (DAMIDL) is proposed for improving AD diagnosis performance, which can automatically capture local and global structural features from sMRI scans and make AD-related classification decisions in a unified framework.

2) The Patch-Nets with spatial attention blocks are designed to extract discriminative features within each patch and to enhance the local features of abnormally changed micro-structures caused by atrophy in the brain.

3) An attention multi-instance learning (MIL) pooling operation is proposed to balance the relative contribution of each patch and yield a global different weighted feature representation for the whole brain structure.

Method

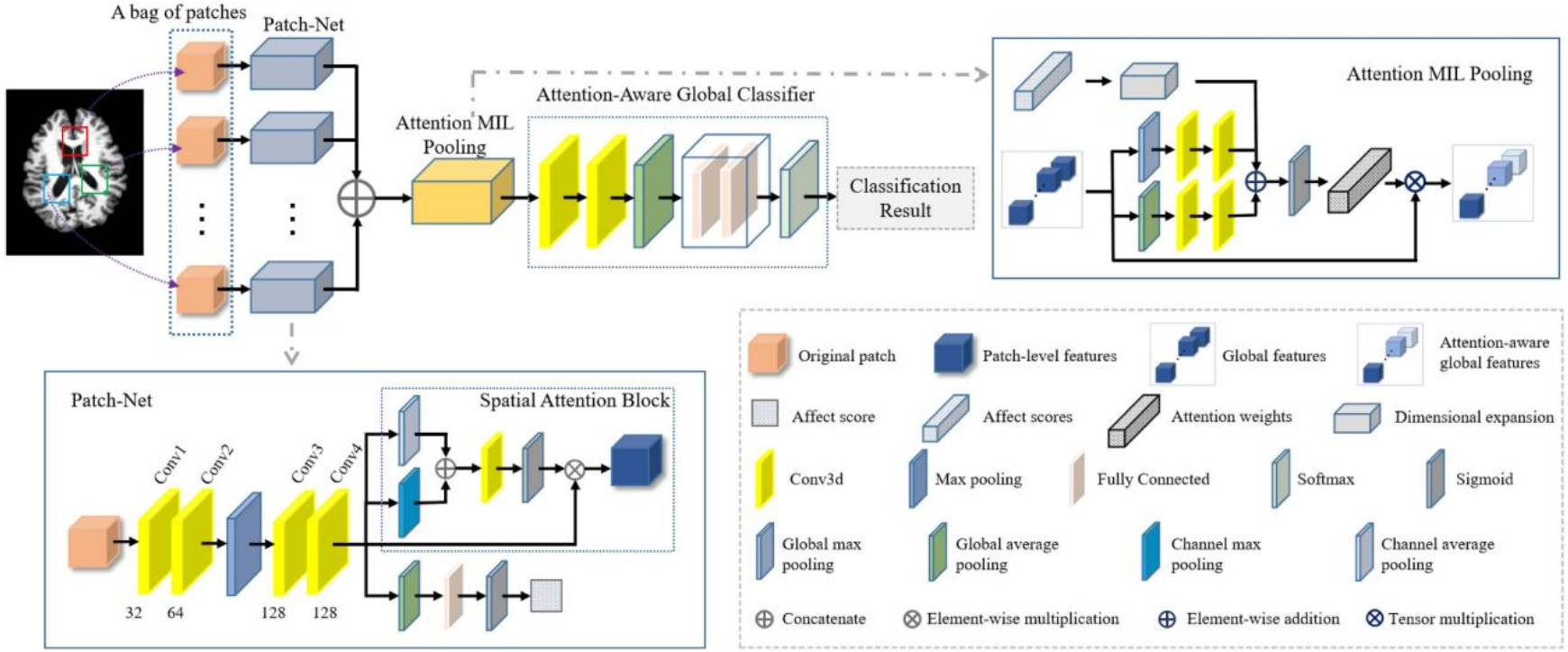


Fig. 1. Illustration of our dual attention multi-instance deep learning network (DA-MIDL), which consists of Patch-Nets with spatial attention blocks, attention MIL pooling and attention-aware global classifier.

1) the **Patch-Nets** with spatial attention blocks for extracting discriminative features within each sMRI patch whilst enhancing the features of abnormally changed micro-structures in the cerebrum,

2) an **attention multi-instance learning (MIL) pooling** operation for balancing the relative contribution of each patch and yield a global different weighted representation for the whole brain structure, and

3) an **attention-aware global classifier** for further learning the integral features and making the AD-related classification decisions.

Method

Patch Location Proposals

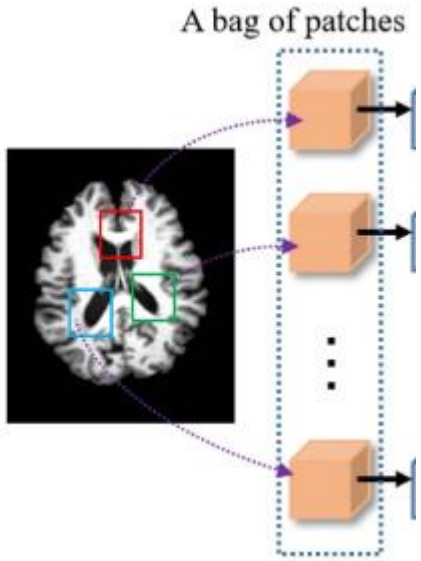
apply the t-tests to sort the informativeness in all patches

(1) calculate the average of the voxel-wise features in one patch as its patch-level feature;

(2) make a group comparison on two groups of patch-level features at one patch location respectively from the same amount of AD patients and normal controls in training set using a t-test;

(3) obtain a p-value at this patch location;

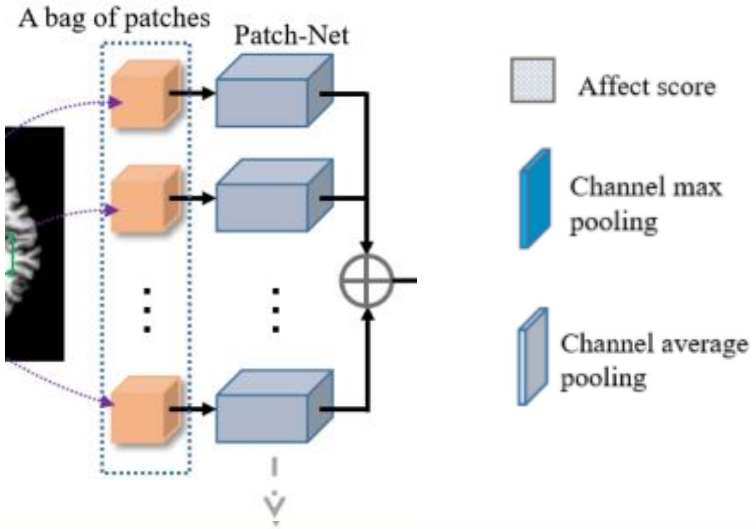
(4) select a number of patches in one image at the locations with the smallest p-values to compose a bag



$$\text{e.g., } \boldsymbol{X} = \{\boldsymbol{I}_1, \boldsymbol{I}_2, \cdots, \boldsymbol{I}_k\}, \text{ where } \boldsymbol{I}_i \in \mathbb{R}^{W \times W \times \hat{W}}$$

Method

Patch-Net With Spatial Attention Block



Spatial Attention Block

1) output of conv4 $\boldsymbol{F} = \{\boldsymbol{F}_1, \boldsymbol{F}_2, \cdots, \boldsymbol{F}_C\}$, $\boldsymbol{F}_i \in \mathbb{R}^{w \times w \times w}$, C is the number of channels

$$\boldsymbol{F}_{max} = ChannelMaxPooling(\boldsymbol{F}),$$

$$\boldsymbol{F}_{average} = ChannelAveragePooling(\boldsymbol{F}),$$

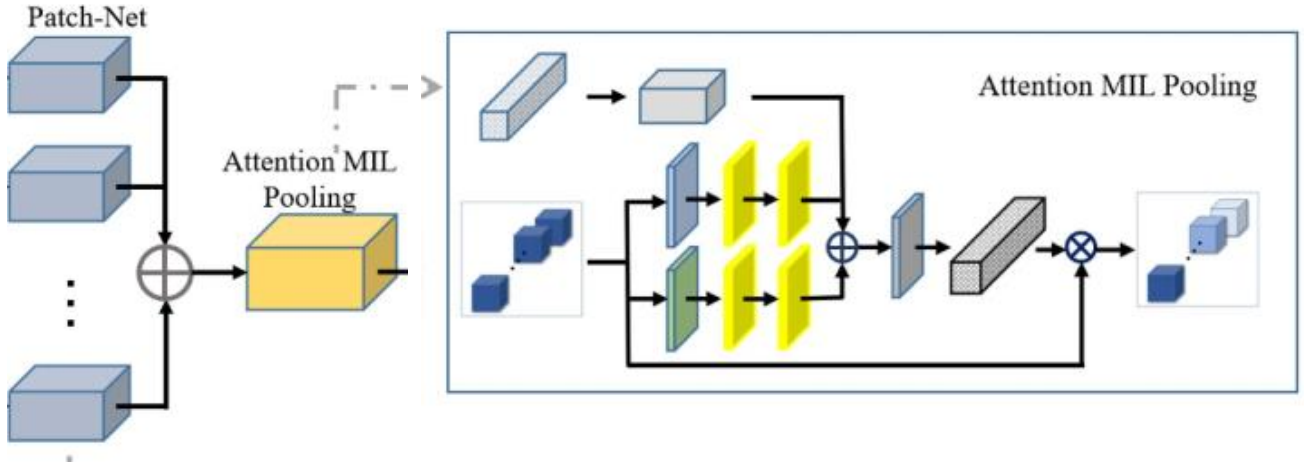
2) concatenate the two feature maps and calculate a spatial attention map

$$\mathbb{A}_{spatial} = \sigma(\mathbb{W}([\boldsymbol{F}_{max}; \boldsymbol{F}_{average}])),$$

3) patch-level spatial-attention-aware feature representation **F**

$$\mathbf{F} = [\boldsymbol{F}_1 \otimes \mathbb{A}_{spatial}; \cdots; \boldsymbol{F}_C \otimes \mathbb{A}_{spatial}],$$

Method



Attention MIL Pooling

1) Input:

   Each patch-level structural representation $\mathbf{F} \in \mathbb{R}^{C\times w\times w\times w}$

2) Compressed by average-pooling along channel axis to $\bar{\mathbf{F}} \in \mathbb{R}^{1\times w\times w\times w}$, Then, the compressed patch-level feature representations are concatenated to the global feature representation as

$$\mathbf{F}_{global} = \{\bar{\mathbf{F}}_1, \bar{\mathbf{F}}_2, \cdots, \bar{\mathbf{F}}_C\}$$

3) The global average pooling (GAP) and global max pooling (GMP) are constructed in parallel for generating two different feature descriptors

$$\boldsymbol{A}_{average} = \mathbb{W}_1 ReLU(\mathbb{W}_0 GAP(\mathbf{F}_{global}))$$
$$\boldsymbol{A}_{max} = \mathbb{W}_1 ReLU(\mathbb{W}_0 GMP(\mathbf{F}_{global}))$$

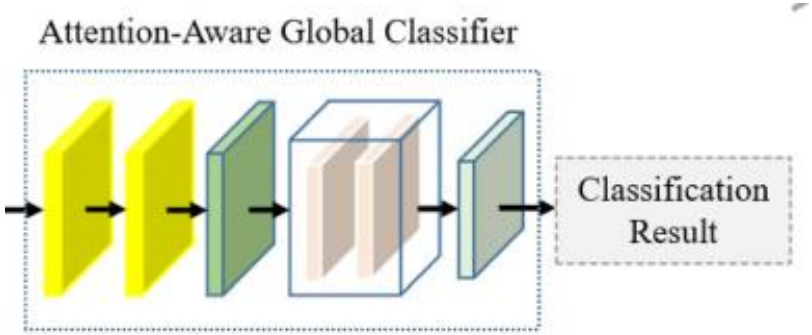
4) the three different attention maps can be merged into a comprehensive patch-attention map $\mathbb{A}_{patch}$ by element-wise summation

$$\mathbb{A}_{patch} = \sigma(\boldsymbol{A}_{average} + \boldsymbol{A}_{max} + \boldsymbol{a}) \qquad \boldsymbol{a} = \{a_1, a_2, \cdots, a_C\}$$

$$\mathcal{F}_{global} = \mathbf{F}_{global} \otimes \mathbb{A}_{patch}$$

Method

Attention-Aware Global Classifier



$$\mathcal{L}(\mathbf{W}) = -\frac{1}{N}\sum_{n=1}^{N} log(P(Y_n|\boldsymbol{X}_n; \mathbf{W})),$$

Experimental Results

Metric:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$SEN = \frac{TP}{TP+FN}$$

$$SPE = \frac{TN}{TN+FP}$$

Experimental Results

TABLE II

RESULTS FOR AD CLASSIFICATION (AD VS. NC) AND MCI CONVERSION PREDICTION (PMCI VS. SMCI) ON THE ADNI TEST SET

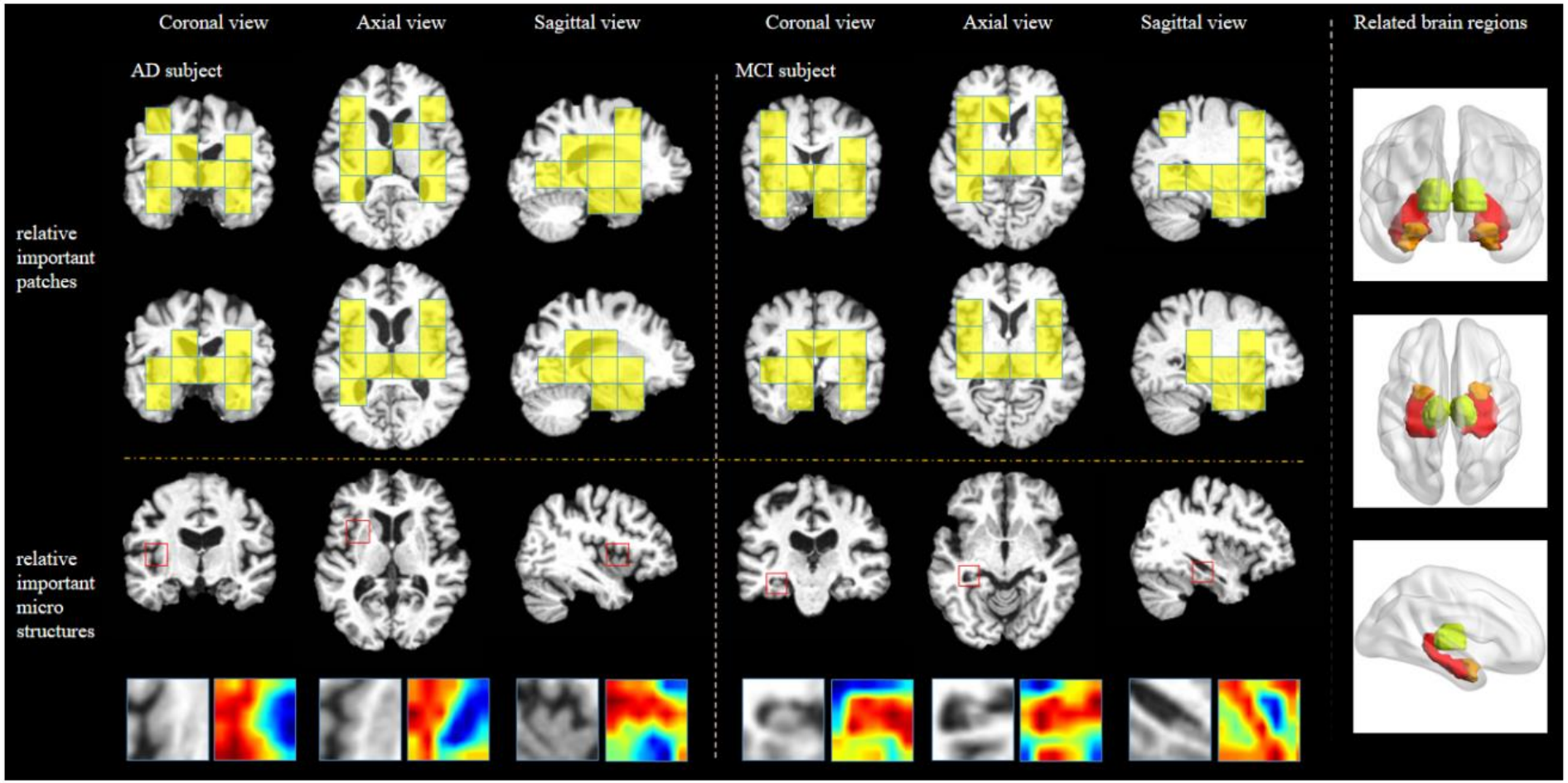
	Method	AD vs. NC				pMCI vs. sMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
voxel level	VBM	0.816	0.756	0.875	0.883	0.679	0.629	0.717	0.709
region level	ROI	0.804	0.718	0.888	0.852	0.667	0.571	0.739	0.692
patch level	PLM	0.848	0.846	0.850	0.905	0.716	0.657	0.761	0.732
	DMIL	0.892	0.859	0.925	0.950	0.765	0.714	0.804	0.790
	HFCN	0.905	0.897	0.913	0.942	0.778	0.686	**0.848**	0.812
	DA-MIDL	**0.924**	**0.910**	**0.938**	**0.965**	**0.802**	**0.771**	0.826	**0.851**

Experimental Results

TABLE III
RESULTS FOR pMCI VS. NC AND sMCI VS. NC CLASSIFICATIONS ON THE ADNI TEST SET

Method	pMCI vs. NC				sMCI vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
VBM	0.816	0.647	0.888	0.853	0.698	0.674	0.713	0.742
ROI	0.789	0.618	0.862	0.846	0.675	0.652	0.688	0.698
PLM	0.825	0.765	0.850	0.876	0.738	0.652	0.788	0.756
DMIL	0.868	0.735	**0.925**	0.908	0.794	0.783	0.800	0.808
HFCN	0.877	0.795	0.913	0.910	0.802	0.717	**0.850**	0.832
DA-MIDL	**0.895**	**0.824**	**0.925**	**0.917**	**0.825**	**0.804**	0.838	**0.860**

Experimental Results



(MICCAI 2021)
Longitudinal Self-supervision to Disentangle Inter-patient Variability from Disease Progression

Rapha¨el Couronn´e(B), Paul Vernhet, and Stanley Durrleman

Inria - Aramis Project-Team, Sorbonne Universit´e, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hˆopital de la Piti´e Salpˆetri`ere, Paris, France

Motivation

1. There is a strong interplay between the pathological progression and the inter-subject variability, which makes it all the more necessary to characterize the contribution of each factor.

2. Typically, in the context of neurodegenerative diseases, we may ask whether the atrophy of a particular brain region is predictive of a specific patient advancement in the disease, or rather can be dismissed as a specific characteristic of the individual.

Contributions

Propose a generic deep longitudinal model, designed to disentangle inter-patient variability from an estimated disease progression timeline.

(i) an architecture that is tailored to disease progression modeling and disentangles the changes due to progression from the changes due to phenotypic differences across subjects;

(ii) a modular method with decoders adapted to data types;

(iii) an application on synthetic and real datasets - including imaging and clinical data - showing that one direction of the latent space alone describes temporal progression.

Methodology

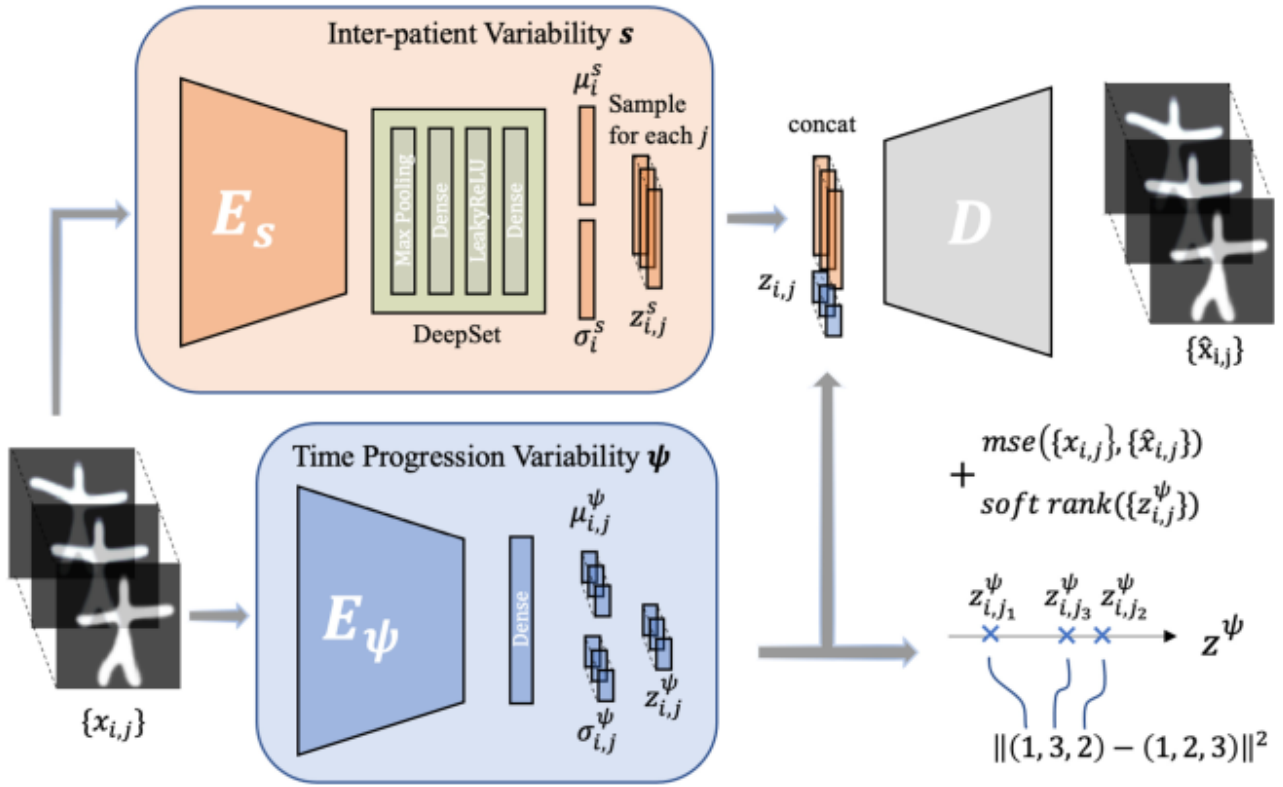
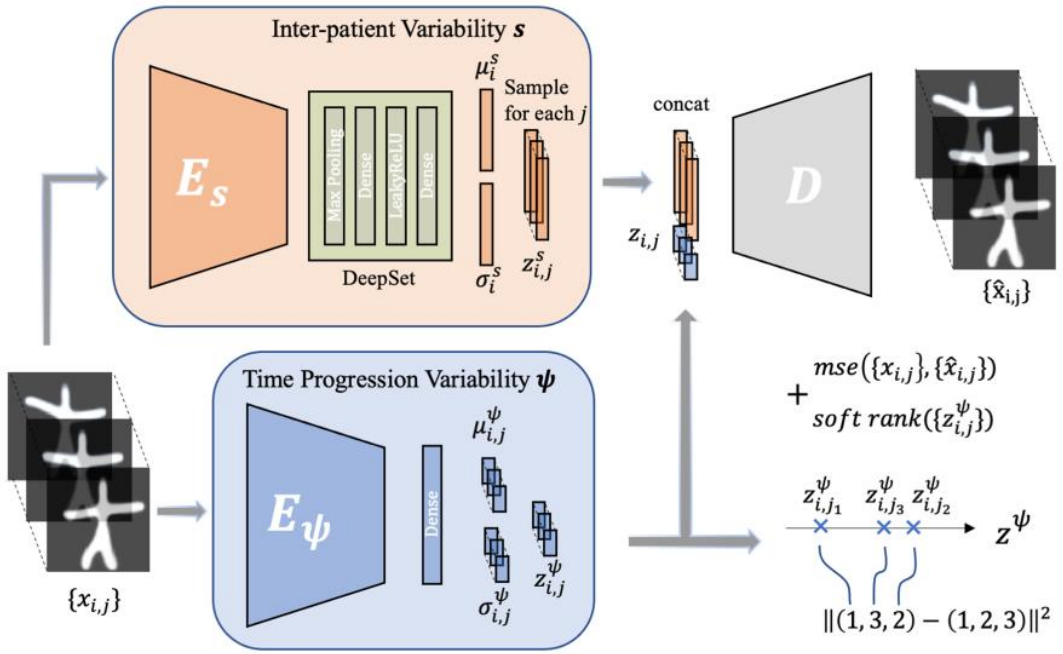


Fig. 1. Input data $\boldsymbol{x_i} = \{x_{i,j}, \forall j \in [1, n_i]\}$ is encoded simultaneously in a space encoder (Deepset) and a point-wise time encoder to get respectively (μ_i^s, σ_i^s) and $(\mu_{i,j}^{\psi}, \sigma_{i,j}^{\psi})$. (μ_i^s, σ_i^s) can be computed from *any* subset of visits, and in practice randomized fixed-size subsets of visits are drawn in the spirit of stochastic optimization. Note that we sample a $z_{i,j}^s$ per visit, but z_i^s could be sampled once for a patient. Decoder can be either agnostic, or specific (e.g., velocity fields for deformations).

Longitudinal Progression Model



Assumption

It assumes that a sequence of observations is generated as the combination of an intrinsic code z^s (as in space shift) and a disease progression factor z^ψ

$$x_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}\left\{\Phi(z_i^s,\ z_{i,j}^\psi);\ \epsilon^2\mathbb{I}\right\} \text{ with } z_i^s \overset{\text{iid}}{\sim} \mathcal{N}(0, \lambda_s^2\mathbb{I}) \text{ and } z_{i,j}^\psi \overset{\text{iid}}{\sim} \mathcal{N}(0, \lambda_\psi^2)$$

Final Objective

$$L = \sum^{N} \mathrm{KL}[q_\eta(\boldsymbol{z_i^\psi}, z_i^s|\boldsymbol{x_i})||p(\boldsymbol{z_i})] - \sum_{j=1}^{n_i} \mathbb{E}[\log p_\theta(x_{i,j}|z_{i,j})] + \gamma \mathcal{C}_i^{\text{ranking}}$$

Results

Validation on Synthetic Data

Table 1. Benchmark of proposed methods on Starmen dataset

Metric	β-VAE	ML-VAE	LR-AE	AR-VAE	LSSL	Ours	Ours (wD)	Ours (woR)
MSE (10^{-3})	7.90 $\pm$ 0.57	22.7 $\pm$ 1.51	10.9 $\pm$ 1.53	8.26 $\pm$ 0.62	7.32 $\pm$ 0.379	8.83 $\pm$ 0.88	**6.22** $\pm$ 1.23	14.2 $\pm$ 5.46
PLS z^{ψ}/z^{s}	- -	0.660 $\pm$ 0.343	0.137 $\pm$ 0.209	0.125 $\pm$ 0.117	0.098 $\pm$ 0.047	**0.083** $\pm$ 0.026	0.083 $\pm$ 0.025	0.149 $\pm$ 0.131
Staging ψ^{*}	0.263 $\pm$ 0.348	0.030 $\pm$ 0.028	0.971 $\pm$ 0.024	0.984 $\pm$ 0.008	0.994 $\pm$ 0.003	**0.997** $\pm$ 0.001	0.996 $\pm$ 0.002	0.524 $\pm$ 0.464

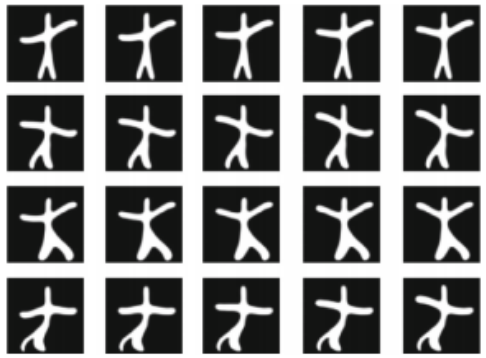


Fig. 2. Each row represents a synthetic subject across time.

Application to Alzheimer's Disease

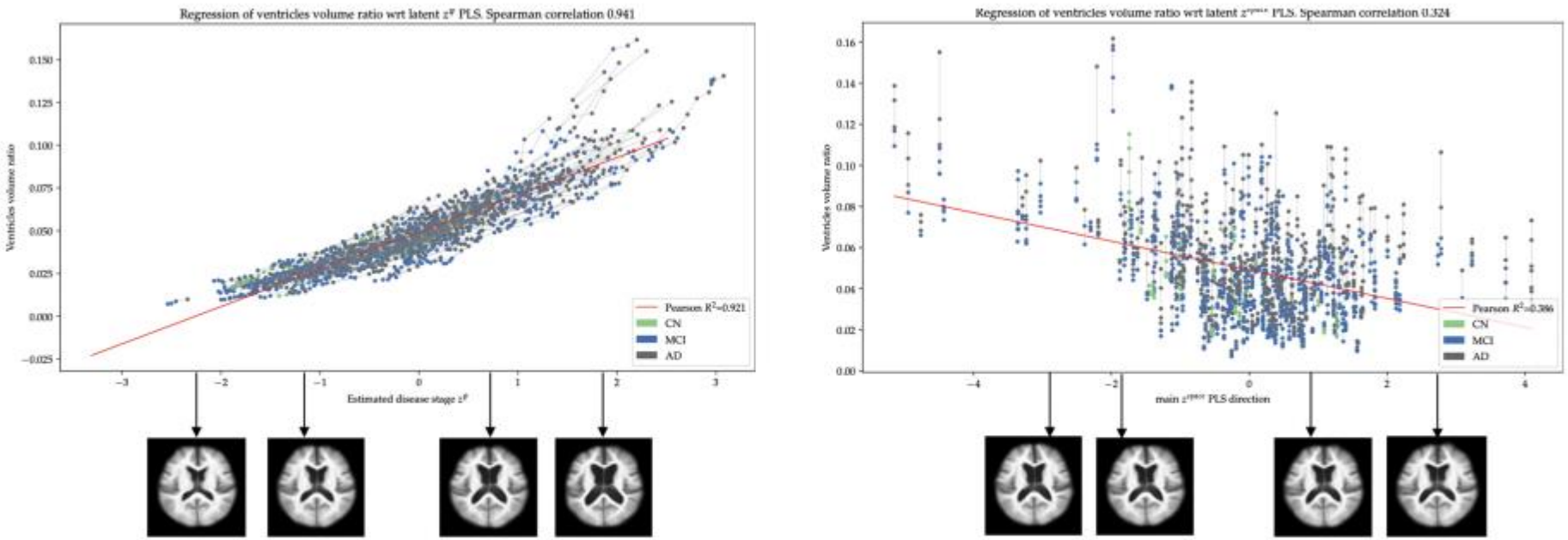


Fig. 5. PLS analysis with respect to ventricle volume ratio $\mathcal{V}$: z^{ψ} (left), z^{s} (right).

Thanks!

2021/12/15

Luoyao Kang