LDGAN: Longitudinal-Diagnostic Generative Adversarial Network for Disease Progression Prediction with Missing Structural MRI

Zhenyuan Ning[1,2], Yu Zhang[1(✉)], Yongsheng Pan[3], Tao Zhong[1,2], Mingxia Liu[2(✉)], and Dinggang Shen[2(✉)]

[1] School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
yuzhang@smu.edu.cn

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mxliu@med.unc.edu, dinggang.Shen@gmail.com

[3] School of Computer Science and Engineering, Northwestern Polytechnical University, Xian 710072, China

Abstract. Predicting future progression of brain disorders is fundamental for effective intervention of pathological cognitive decline. Structural MRI provides a non-invasive solution to examine brain pathology and has been widely used for longitudinal analysis of brain disorders. Previous studies typically use only *complete baseline* MRI scans to predict future disease status due to the lack of MRI data at one or more future time points. Since temporal changes of each brain MRI are ignored, these methods would result in sub-optimal performance. To this end, we propose a longitudinal-diagnostic generative adversarial network (LDGAN) to predict multiple clinical scores at future time points using incomplete longitudinal MRI data. Specifically, LDGAN imputes MR images by learning a bi-directional mapping between MRIs of two adjacent time points and performing clinical score prediction jointly, thereby explicitly encouraging task-oriented image synthesis. The proposed LDGAN is further armed with a *temporal constraint* and an *output constraint* to model the temporal regularity of MRIs at adjacent time points and encourage the diagnostic consistency, respectively. We also design a weighted loss function to make use of those subjects without ground-truth scores at certain time points. The major advantage of the proposed LDGAN is that it can impute those missing scans in a task-oriented manner and can explicitly capture the temporal characteristics of brain changes for accurate prediction. Experimental results on both ADNI-1 and ADNI-2 datasets demonstrate that, compared with the state-of-the-art methods, LDGAN can generate more reasonable MRI scans and efficiently predict longitudinal clinical measures.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59861-7_18) contains supplementary material, which is available to authorized users.

M. Liu et al. (Eds.): MLMI 2020, LNCS 12436, pp. 170–179, 2020.
https://doi.org/10.1007/978-3-030-59861-7_18

1 Introduction

Structural magnetic resonance imaging (MRI) provides a feasible solution to potentially identify abnormal changes of the brain that could be used as biomarkers for automated diagnosis of brain diseases, such as Alzheimer's Disease (AD) and its prodromal stage (i.e., mild cognitive impairment, MCI) [1–5]. An interesting topic is to assess the stage of pathology and predict future progression of MCI by estimating longitudinal clinical scores of subjects based on MRI data [6,7]. However, missing data problem (e.g., incomplete MRI and ground-truth clinical scores) has been remaining a huge challenge for longitudinal diagnosis of brain diseases due to patient dropout and/or poor data quality.

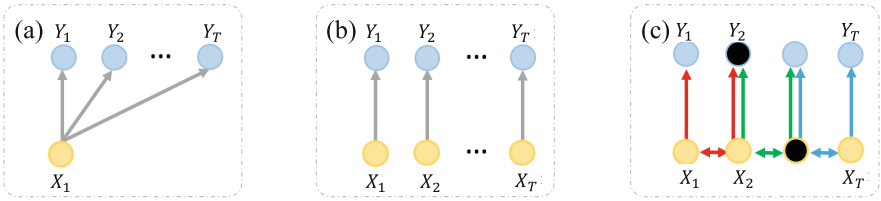

Fig. 1. Illustration of three learning strategies for MRI-based longitudinal diagnosis: (a) estimating baseline and T future clinical scores (i.e., $\{Y_1, Y_2, \cdots, Y_T\}$) using only baseline MRI (i.e., X_1); (b) estimating each score using its corresponding true MRI at the same time point; and (c) estimating scores at multiple time points using incomplete longitudinal MRI, where those missing MRIs (denoted as black circles) are imputed by a certain image synthesis method. Two-way arrows and one-way arrows represent image synthesis and diagnosis, respectively.

Existing learning-based methods often simply discard subjects without MRI or ground-truth scores at a certain time point, which results in limited training samples for learning reliable and robust models [8,9]. Recent effort has been devoted to taking advantage of all available subjects by using multi-view learning or data imputation techniques [2,7]. For example, a weakly-supervised densely connected neural network (WiseDNN) is proposed to perform longitudinal diagnosis for AD and MCI based on *complete baseline* MRI and incomplete clinical scores, as shown in Fig. 1 (a). However, these methods generally ignore temporal changes of brain MRIs and clinical scores, because only baseline MRIs are used to predict future clinical measures. It seems to be more reliable to predict clinical scores using MRIs at the same time point, as shown in Fig. 1 (b). On the other hand, even though several studies propose to directly impute missing data [2,10,11], they typically treat data imputation and disease diagnosis as two standalone tasks. Intuitively, integrating MRI synthesis and disease diagnosis into a unified framework is desirable, as illustrated in Fig. 1 (c).

In this work, we propose a longitudinal-diagnostic generative adversarial network (LDGAN) for joint longitudinal image synthesis and clinical score prediction based on incomplete MRIs and ground-truth clinical scores. As shown in

Fig. 2, the proposed LDGAN can jointly perform image synthesis and disease diagnosis at multiple (e.g., T) time points. Specifically, by using the current time-point MR images (e.g., X_t) as input, LDGAN is developed to synthesize MR images (e.g., $\tilde{X}_{t+1}$) at the next time point, through which imaging representations that reflect temporal changes of the brain over time can be simultaneously learned for diagnosis. Moreover, our LDGAN is armed with a temporal constraint (e.g., R_t^T) and an output constraint (e.g., R_t^C) to model the temporal regularity and encourage the diagnostic consistency, respectively. To make full use of all available subjects (even those without ground-true scores at a certain time point), a weighted loss function is further employed to train the network. Experimental results on ADNI-1 and ADNI-2 demonstrate that the proposed method can simultaneously generate reasonable MRI scans and efficiently predict longitudinal clinical scores, compared with several state-of-the-art methods.

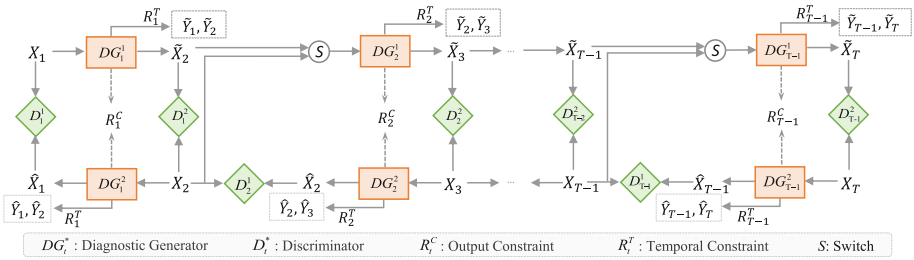

Fig. 2. Overview of the proposed LDGAN for joint image synthesis and clinical score prediction at multiple (i.e., T) time points based on incomplete MRI data, including 1) diagnostic generator (e.g., DG_1^1 and DG_1^2) for image synthesis, 2) discriminator (e.g., D_1^1 and D_1^2), and 3) temporal constraint (e.g., R_1^T) and output constraint (e.g., R_1^C) . The switch (S) operation indicates that the input is a real image if it exists; otherwise, the input is a synthetic image.

2 Method

2.1 Problem Analysis

For the i-th subject, we denote $\{X_t^i, Y_t^i\}_{t=1}^T$ as its MRI scans and clinical scores at T time points. To avoid the time gap problem, we use MRI scans at the t-th time point to predict clinical scores at the t-th time point. Specifically, the longitudinal diagnostic model can be formulated as $\tilde{Y}_t^i = \Phi_t(X_t^i)$, where Φ_t and $\tilde{Y}_t^i$ are the regressor and the predicted score for i-th subject at the t-th time point, respectively. However, the predictor $\Phi_t(-)$ cannot be executed if X_t^i is missing. To address this issue, the straightforward way is to use the real MRI of i-th subject at another time point to generate a virtual X_t^i. Considering that AD/MCI are progressive neurodegenerative disorders, we assume that a missing

MRI scan at the next time point (e.g., X_{t+1}^i) can be synthesized by the real MRI scan at current time point (e.g., X_t^i), as shown in Fig. 1 (c).

Let G_t be the function to generate X_{t+1}^i from X_t^i, and D_t as the discriminator to distinguish whether the image is real or synthetic. Based on synthetic images, a diagnostic model at the t-th time point can be formulated as $\tilde{Y}_t^i \approx \Phi_t(G_t(X_t^i))$. Previous studies have shown that temporal changes of the brain are beneficial to longitudinal diagnosis [12–14]. Accordingly, we assume that MRI features that reflect the brain changes of can be jointly learned for diagnosis during the process of image generation, and denote such a joint image synthesis and diagnosis model as DG_t. With some constraint terms (denoted as $R(-)$), the longitudinal-diagnostic model with can be reformulated as

$$\tilde{Y}_t^i = \Phi_t(X_t^i) \approx DG_t(X_t^i), \quad s.t.\ R(DG_t, D_t). \tag{1}$$

To this end, we propose a longitudinal-diagnostic generative adversarial network (LDGAN) to jointly perform image synthesis and diagnosis at multiple time points, with the schematic illustration shown in Fig. 2. In the following, we first introduce LDGAN for joint image synthesis and diagnosis at two time points, and then extend it for the task with multiple time points.

2.2 LDGAN with Two Time Points

The architecture of LDGAN at two time points is illustrated in Fig. 3, which basically consists of diagnostic generator (e.g., DG_t), discriminator (e.g., D_t), and constraint operator (e.g., R_t^T and R_t^C), with the details given below.

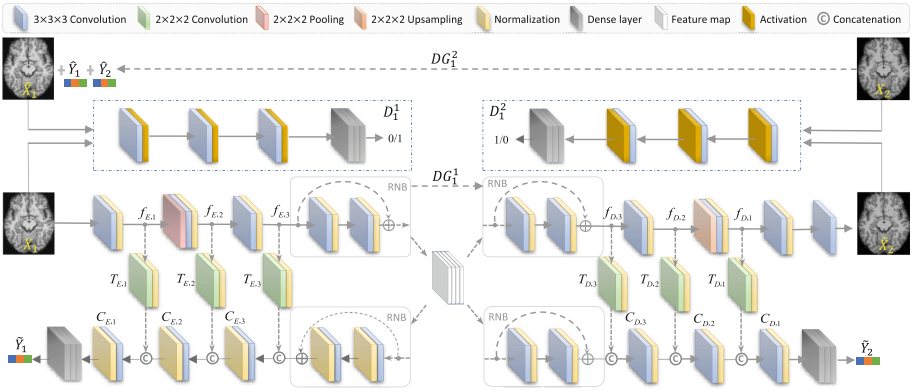

Fig. 3. Network architecture of LDGAN for two time-point image synthesis and clinical score prediction, including two longitudinal-diagnostic generators (e.g., DG_1^1 with X_1 as input and $\{\tilde{X}_2, \tilde{Y}_1, \tilde{Y}_2\}$ as output, and DG_1^2 with X_2 as input and $\{\hat{X}_1, \hat{Y}_1, \hat{Y}_2\}$ as output), and two discriminators (e.g., D_1^1 and D_1^2). RNB: residual network block.

Diagnostic Generator. The two time-point LDGAN has two diagnostic generators, i.e., $DG_1^1 : X_1^i \rightarrow \{\tilde{X}_2^i, \tilde{Y}_1^i, \tilde{Y}_2^i\}$ for the 1st time point and $DG_1^2 : X_2^i \rightarrow \{\hat{X}_1^i, \hat{Y}_1^i, \hat{Y}_2^i\}$ for the 2nd time point. Each diagnostic generator contains a shared encoder, a decoder, and two predictors, for joint image synthesis and prediction.

1) Shared Encoder. The shared encoder aims to extract source-domain-specific features and transferable features shared by source and target domains. Specifically, the shared encoder consists of a set of convolutional blocks and a residual network block (RNB). Three consecutive convolutional blocks are first used to extract MRI features, each of which containing a convolutional layer (with the kernel size of $3 \times 3 \times 3$) and a normalization layer. Between the first two blocks, a $2 \times 2 \times 2$ pooling layer is placed to perform feature reduction. The subsequent RNB consists of two convolutional blocks with the dropout rate of 0.8 and skip-connections. The RNB is used to generate shared feature maps for the source domain (e.g., $\mathcal{X}_1$) and the target domain (e.g., $\mathcal{X}_2$). And such feature maps are further used for image synthesis (via the decoder) and clinical score prediction (via two predictors).

2) Decoder. The decoder is constructed to learn target-domain-specific features and produce synthesized images based on feature maps produced by the encoder. The structure of the decoder is basically mirrored from the shared encoder, i.e., with an RNB followed by three convolutional blocks. In particular, the connective order of decoder is opposite to encoder and the pooling layer is replaced by an upsampling layer. A convolutional layer is further used to generate MRI scans in the target domain.

3) Predictor. Unlike previous predictors that are independent of image generation, the predictors in our LDGAN are embedded in the diagnostic generators. Each predictor basically contains an RNB, three convolutional blocks (denoted as $\{C_{E,j}\}_{j=1}^3$ and $\{C_{D,j}\}_{j=1}^3$ for predictors at the current time point and the next time point, respectively), and three dense layers.

The transferable/shared features learned from the encoder are first fed into an RNB to extract high-level semantic features. Since the domain-specific features can provide disease-related information, each predictor adopts a hierarchical fusion strategy to integrate the shared features and domain-specific features for diagnosis. As shown in Fig. 3, we denote the domain-specific features generated by j-th convolutional block in encoder and decoder as $f_{E,j}$ and $f_{D,j}$, respectively. The feature map $f_{E,j}$ (or $f_{D,j}$) is fed into a transfer block (i.e., $T_{E,j}$ or $T_{D,j}$) that consists of a $2 \times 2 \times 2$ convolutional layer and a normalization layer to generate the j-th level domain-specific features. These features are further concatenated with the output of the convolutional block (i.e., $C_{E,j+1}$ or $C_{D,j+1}$) or the residual block in each predictor. Then, the concatenated features are fed into the convolutional block (i.e., $C_{E,j}$ or $C_{D,j}$), whose output is concatenated with the (j-1)-level domain-specific features as the input of the next convolutional block (i.e., $C_{E,j-1}$ or $C_{D,j-1}$). Subsequently, a global pooling layer and three dense layers are used to predict the scores based on the features generated by the last convolutional block (i.e., $C_{E,1}$ or $C_{D,1}$). The loss function

of the proposed diagnostic generator (with DG_1^1 for the current time point and DG_1^2 for the next time point) is defined as:

$$\mathcal{L}(DG_1^1, DG_1^2) = \mathcal{L}_{DG}(DG_1^1) + \mathcal{L}_{DG}(DG_1^2) + \lambda_c \mathcal{L}_C(DG_1^1, DG_1^2), \tag{2}$$

$$\begin{aligned}\mathcal{L}_{DG}(DG_1^1) = &\mathbb{E}_{X_1^i \in \mathcal{X}_1} ||\log(D_1^2(G_1^1(X_1^i)))||_2 + \mathbb{E}_{X_1^i \in \mathcal{X}_1} ||P_1^{1,1}(X_1^i) - Y_1^i||_2 \\ &+ \mathbb{E}_{X_1^i \in \mathcal{X}_1} ||P_1^{1,2}(X_1^i) - Y_2^i||_2,\end{aligned} \tag{3}$$

where G_1^1 is the generator, and $P_1^{1,1}$ and $P_1^{1,2}$ are the predictors for the 1st and 2nd time points in DG_1^1, respectively. Using Eq. (3), one can ensure that the synthetic images can be generated in a prediction-oriented manner by using both domain-specific features and shared features. Besides, the last term in Eq. (2) is defined as:

$$\mathcal{L}_{\mathcal{C}}(DG_1^1, DG_1^2) = \mathbb{E}_{X_1^i \in \mathcal{X}_1} ||G_1^2(G_1^1(X_1^i)) - X_1^i||_2 + \mathbb{E}_{X_2^i \in \mathcal{X}_2} ||G_1^1(G_1^2(X_2^i)) - X_2^i||_2, \tag{4}$$

which is used to encourage each synthetic MR image to be consistent with its corresponding real MR image.

Discriminator. The proposed LDGAN contains two adversarial discriminators, i.e., D_1^1 and D_1^2, to distinguish whether the image is real or synthetic. Specifically, each discriminator contains three consecutive convolutional layers with LeakyReLU activation function and kernel size of $3 \times 3 \times 3$ to extract imaging features. By leveraging these features, three dense layers (with the neuron number of 16, 8, and 1, respectively) are used to perform identification. The loss function of the discriminator is defined as:

$$\begin{aligned}\mathcal{L}(D_1^1, D_1^2) = &\mathbb{E}_{X_1^i \in \mathcal{X}_1} ||\log(D_1^1(X_1^i))||_2 + \mathbb{E}_{X_1^i \in \mathcal{X}_1} ||\log(1 - D_1^1(G_1^2(X_2^i)))||_2 \\ &+ \mathbb{E}_{X_2^i \in \mathcal{X}_2} ||\log(D_1^2(X_2^i))||_2 + \mathbb{E}_{X_2^i \in \mathcal{X}_2} ||\log(1 - D_1^2(G_1^1(X_1^i)))||_2.\end{aligned} \tag{5}$$

Constraints. Our LDGAN includes a temporal constraint and an output constraint, both of which are used to constrain the output of two predictors. Specifically, the proposed temporal and output constraints are defined as:

$$R(DG_1^1, DG_1^2) = R^T(DG_1^1, DG_1^2) + \mu_c R^C(DG_1^1, DG_1^2), \tag{6}$$

where μ_c is the penalty coefficient, and the temporal constraint is defined as:

$$R^T(DG_1^1, DG_1^2) = \mathbb{E}_{X_1^i \in \mathcal{X}_1} \psi(\tilde{Y}_1^i, \tilde{Y}_2^i) + \mathbb{E}_{X_2^i \in \mathcal{X}_2} \psi(\hat{Y}_1^i, \hat{Y}_2^i), \tag{7}$$

where the term $\psi(A, B) = 0$ if A and B follows the time regularity of clinical scores (e.g., longitudinal increase in Clinical Dementia Rating and longitudinal decrease in Alzheimer's Disease Assessment Scale's Cognitive subscale); and $\psi(A, B) = ||A\text{-}B||_2$, otherwise. Equation 7 encourages the predicted scores to follow time regularity of clinical scores. The output constraint is defined as:

$$R^C(DG_1^1, DG_1^2) = \mathbb{E}_{X_1^i \in \mathcal{X}_1, X_2^i \in \mathcal{X}_2} ||\tilde{Y}_1^i - \hat{Y}_1^i||_2 + \mathbb{E}_{X_1^i \in \mathcal{X}_1, X_2^i \in \mathcal{X}_2} ||\tilde{Y}_2^i - \hat{Y}_2^i||_2, \tag{8}$$

which encourages the outputs of predictors for the same score at the same time point to be consistent.

2.3 LDGAN with Multiple Time Points

The LDGAN can be further extended to handle problems with multiple time points. As shown in Fig. 2, DG_1 is first developed to map X_1 to $\tilde{X}_2$, and predict the scores $\tilde{Y}_1$ and $\tilde{Y}_2$. Then, $\tilde{X}_2$ is used to complement X_2 as the input of DG_2 to synthesize $\tilde{X}_3$ and predict the scores $\tilde{Y}_2$ and $\tilde{Y}_3$. By that analogy, we can generate the missing images and its clinical scores. To make use of subjects without ground-truth clinical scores, a weighted loss function is designed in LDGAN to make use of all available subjects (even those with missing ground-true scores at a certain time point). Let $Y_t = [\mathbf{y}_{t,1}, \cdots, \mathbf{y}_{t,s}, \cdots, \mathbf{y}_{t,S}]$ denote S types of ground-truth clinical scores of all subjects at the t-th time point. The $\mathcal{L}_{DG}(DG_t^1)$ and $\mathcal{L}_{DG}(DG_t^2)$ in $\mathcal{L}(DG_t^1, DG_t^2)$ are written as:

$$\begin{aligned}\mathcal{L}_{DG}(DG_t^*) = &\mathbb{E}_{X_t^i \in \mathcal{X}_t} ||\log(1 - D_t^2(G_t^*(X_t^i)))||_2 + \mathbb{E}_{X_t^i \in \mathcal{X}_t} \mathbf{H}_t^i \times ||P_t^{*,1}(X_t^i) - Y_t^i||_2 \\ &+ \mathbb{E}_{X_t^i \in \mathcal{X}_t} \mathbf{H}_{t+1}^i \times ||P_t^{*,2}(X_t^i) - Y_{t+1}^i||_2,\end{aligned} \tag{9}$$

where $\mathbf{H}_t = [\gamma_{t,1}, \cdots, \gamma_{t,s}, \cdots, \gamma_{t,S}]$ is an indicator matrix that denotes whether X_t is labeled with clinical scores or not. To be specific, $\gamma_{t,s}^i = 1$ if the ground-true score $y_{t,s}^i$ is available for X_t^i; and $\gamma_{t,s}^i = 0$, otherwise. Finally, the adversarial loss of LDGAN with multiple time points is defined as:

$$\mathcal{L} = \sum_{t=1}^{T} \mathcal{L}(DG_t^1, DG_t^2) + \mathcal{L}(D_t^1, D_t^2) + \eta_t R(DG_t^1, DG_t^2), \tag{10}$$

where η_t is the penalty coefficient for the t-th time point. In the implementation, we first train D_t and DG_t with fixed predictors, iteratively. After that, we jointly train the generator and predictor in DG_t. The Adam solver is used as the optimizer, with a batch size of 2 and a learning rate of 0.001.

3 Experiment

Data and Experimental Setup. We evaluated LDGAN on subjects with longitudinal MRI from ADNI [15], including ADNI-1 and ADNI-2. Our goal is to predict three types of clinical scores at four time points, i.e., Baseline, the 6th month (M06), 12th month (M12), and 24th month (M24) after baseline. These clinical sores include: 1) clinical dementia rating sum of boxes (CDR-SB), 2) classic AD assessment scale cognitive subscale with 11 items (ADAS-Cog11), and 3) modified ADAS-Cog with 13 items (ADAS-Cog13). After removing subjects that exist in both ADNI-1 and ADNI-2 from ADNI-2, we obtain a total of 824 and 637 subjects from ADNI-1 and ADNI-2, respectively. All studied subjects have baseline MRI data, and only a part of them have M06/M12/M24 MRI data. The number of MRIs and clinical scores of studied subjects are summarized in the *Supplementary Materials*. For all MRI scans, we performed skull-stripping, intensity correction and spatial normalization. Hence, there is spatial correspondence between a set of MRIs at different time points for each subject. ADNI-1 and ADNI-2 are used as training and test sets, respectively.

Performance of MRI Generation. We first evaluate the quality of synthetic MR images generated by our LDGAN and two classical GAN models, i.e., 1) a conventional GAN [16], and 2) the cycle-consistent GAN (CGAN) [17]. It's worthy mentioning that the structures of GAN and CGAN are modified for 3D-input version based on [16] and [17]. Three metrics are used to measure the quality of generated images, including the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) [18]. All models are trained and tested based on the same datasets, namely, subjects in ADNI-1 for training, and subjects in ADNI-2 for testing. The quantitative results are listed in Table 1, from which one can observe that our LDGAN consistently outperforms the competing methods in terms of three metrics at all four time points. We further visually show real and synthetic MR images of a randomly-selected subject from ADNI-2 at four time points in Fig. 4. As can be seen from Fig. 4, our synthetic MR images look more similar to their corresponding real images when compared with other two models. These results demonstrate that our method can generate reasonable longitudinal MRI scans.

Table 1. Comparison of different methods in longitudinal MR image generation at four time points, in terms of MAE (%), SSIM (%), and PSNR.

Method	Baseline			M06			M12			M24		
	MAE	SSIM	PSNR	MAE	SSIM	PSNR	MAE	SSIM	PSNR	MAE	SSIM	PSNR
GAN	4.17	59.38	23.46	4.89	57.66	22.77	4.75	54.22	21.52	8.93	50.77	20.47
CGAN	3.78	62.43	25.83	4.33	58.75	25.62	4.59	54.72	25.04	8.24	51.49	21.42
LDGAN (Ours)	**2.45**	**66.65**	**28.82**	**3.14**	**60.81**	**27.19**	**3.53**	**56.91**	**26.42**	**6.39**	**54.04**	**21.60**

Table 2. Comparison of different methods in predicting three types of clinical scores at four time points using the same ADNI-1 and ADNI-2 datasets, in terms of RMSE.

Method	Baseline			M06		
	CDR-SB	ADAS-Cog11	ADAS-Cog13	CDR-SB	ADAS-Cog11	ADAS-Cog13
LMF [19]	1.922	5.835	8.286	2.394	7.640	10.060
WiseDNN [7]	1.619	5.662	**7.596**	2.016	6.238	8.649
LDGAN (Ours)	**1.572**	**4.980**	7.739	**1.986**	**6.008**	**8.380**
Method	M12			M24		
	CDR-SB	ADAS-Cog11	ADAS-Cog13	CDR-SB	ADAS-Cog11	ADAS-Cog13
LMF [19]	2.694	8.140	10.060	4.009	11.145	14.324
WiseDNN [7]	2.442	7.300	9.888	3.412	9.410	11.177
LDGAN (Ours)	**1.859**	**6.014**	**8.609**	**2.077**	**8.160**	**8.502**

Performance of Longitudinal Diagnosis. We further evaluate the performance of the proposed LDGAN in predicting three types of clinical scores (CDR-SB, ADAS-Cog11, and ADAS-Cog13) at four time points. We compare our model with two state-of-the-art methods, including 1) a model using landmark-based morphological features (LMF) [19], and 2) WiseDNN [7]. Note that LMF and WiseDNN use only baseline MRIs for longitudinal prediction, while our LDGAN

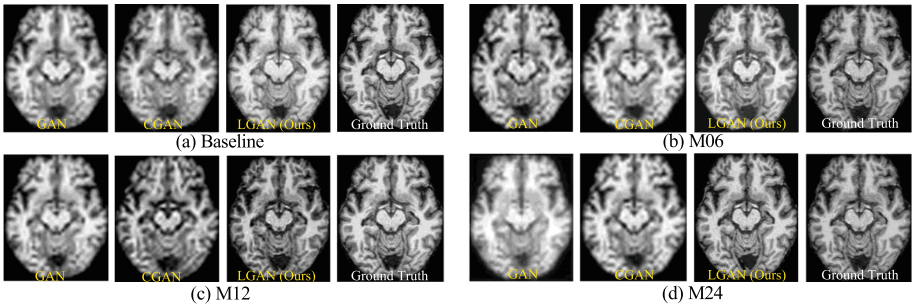

Fig. 4. Illustration of synthetic images generated by three methods and their corresponding ground-truth images at four time points for a subject from ADNI-2.

can use all available MRIs at multiple time points. For a fair comparison, these three methods are trained on ADNI-1 and tested on ADNI-2. The root mean square error (RMSE) is used to evaluate the effectiveness of all methods, with results reported in Table 2. From Table 2, we can observe that, in most cases, the proposed LDGAN obtains the best performance when compared with two competing methods, implying that generating prediction-oriented MRIs (as we do in LDGAN) helps promote the performance of longitudinal prediction.

4 Conclusion

In this work, we propose a longitudinal-diagnostic generative adversarial network (LDGAN) to predict multiple clinical scores at future time points using incomplete longitudinal MRI data. Specifically, LDGAN imputes MR images by jointly learning a bi-directional mapping between MRIs of two adjacent time points and performing clinical score prediction. In addition, the proposed LDGAN is armed with a temporal constraint and an output constraint to encourage the temporal consistency of MRIs at adjacent time points and the output consistency of predicted clinical scores, respectively. To make use of subjects without ground-truth clinical scores at a certain time point, we further design a weighted loss function to train LDGAN. Experimental results demonstrate that LDGAN can generate more reasonable MRI scans and efficiently predict longitudinal clinical scores.

References

1. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. **6**(2), 67–77 (2010)
2. Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J.: Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. NeuroImage **61**(3), 622–632 (2012)
3. Liu, M., Zhang, D., Shen, D.: Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. IEEE Trans. Med. Imaging **35**(6), 1463–1474 (2016)

4. Cheng, B., Liu, M., Shen, D., Li, Z., Zhang, D.: Multi-domain transfer learning for early diagnosis of Alzheimer's disease. Neuroinformatics **15**(2), 115–132 (2017)
5. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
6. Lei, B., Jiang, F., Chen, S., Ni, D., Wang, T.: Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning. Front. Aging Neurosci. **9**(6), 1–6 (2017)
7. Liu, M., Zhang, J., Lian, C., Shen, D.: Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores. IEEE Trans. Cybern. **50**(7), 3381–3392 (2020)
8. Weiner, M.W., et al.: The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer's Dement. **9**(5), e111–e194 (2013)
9. Jiang, P., Wang, X., Li, Q., Jin, L., Li, S.: Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of Alzheimer's disease. IEEE J. Biomed. Health Inform. **23**(4), 1450–1456 (2018)
10. Ritter, K., Schumacher, J., Weygandt, M., Buchert, R., Allefeld, C., Haynes, J.D.: Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. Alzheimer's Dement.: Diagn. Assess. Dis. Monit. **1**(2), 206–215 (2015)
11. Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D.: Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 455–463. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_52
12. Beason-Held, L.L., O Goh, J., An, Y., A Kraut, M., J O'Brien, R., Ferrucci, L.: A longitudinal study of brain anatomy changes preceding dementia in down syndromechanges in brain function occur years before the onset of cognitive impairment. J. Neurosci. **33**(46), 18008–18014 (2013)
13. Jie, B., Liu, M., Liu, J., Zhang, D., Shen, D.: Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease. IEEE Trans. Biomed. Eng. **64**(1), 238–249 (2016)
14. Pujol, J., et al.: A longitudinal study of brain anatomy changes preceding dementia in down syndrome. NeuroImage Clin. **18**, 160–166 (2018)
15. Jack, C., Bernstein, M., Fox, N., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging **27**(4), 685–691 (2008)
16. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
17. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
18. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: International Conference on Pattern Recognition, pp. 2366–2369. IEEE (2010)
19. Zhang, J., Liu, M., An, L., Gao, Y., Shen, D.: Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. IEEE J. Biomed. Health Inform. **21**(6), 1607–1616 (2017)