



Disease-Image Specific Generative Adversarial Network for Brain Disease Diagnosis with Incomplete Multi-modal Neuroimages

Yongsheng Pan^{1,2}, Mingxia Liu²(✉), Chunfeng Lian², Yong Xia¹(✉),
and Dinggang Shen²(✉)

¹ School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China

yxia@nwpu.edu.cn

² Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

mxliu@unc.edu, dgshen@med.unc.edu

Abstract. Incomplete data problem is unavoidable in automated brain disease diagnosis using multi-modal neuroimages (e.g., MRI and PET). To utilize all available subjects to train diagnostic models, deep networks have been proposed to directly impute missing neuroimages by treating all voxels in a 3D volume equally. These methods are not diagnosis-oriented, as they ignore the *disease-image specific information* conveyed in multi-modal neuroimages, i.e., (1) disease may cause abnormalities only at local brain regions, and (2) different modalities may highlight different disease-associated regions. In this paper, we propose a unified disease-image specific deep learning framework for joint image synthesis and disease diagnosis using incomplete multi-modal neuroimaging data. Specifically, by using the whole-brain images as input, we design a disease-image specific neural network (DSNN) to implicitly model disease-image specificity in MRI/PET scans using the spatial cosine kernel. Moreover, we develop a feature-consistent generative adversarial network (FGAN) to synthesize missing images, encouraging DSNN feature maps of synthetic images and their respective real images to be consistent. Our DSNN and FGAN can be jointly trained, by which missing images are imputed in a task-oriented manner for brain disease diagnosis. Experimental results on 1,466 subjects suggest that our method *not only* generates reasonable neuroimages, *but also* achieves the state-of-the-art performance in both tasks of Alzheimer's disease (AD) identification and mild cognitive impairment (MCI) conversion prediction.

1 Introduction

Multi-modal neuroimaging, such as structural magnetic resonance imaging (MRI) and positron emission tomography (PET), provides complementary information and have been widely used in computer-aided diagnosis of Alzheimer's

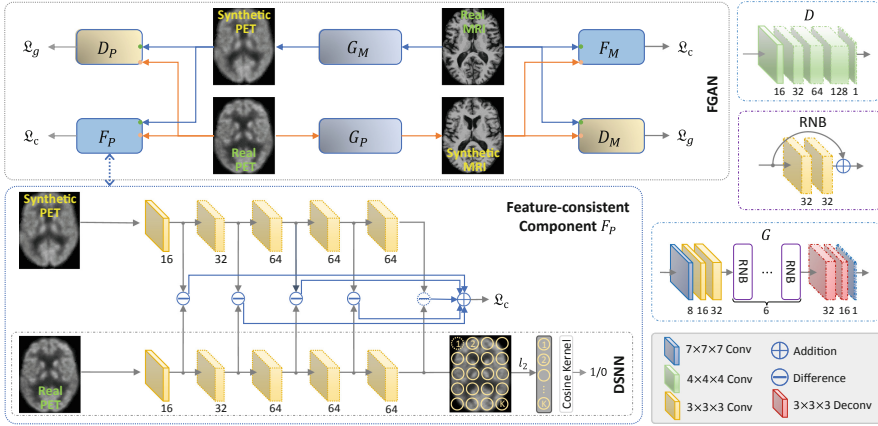


Fig. 1. Illustration of our disease-image specific deep learning framework, including (1) a disease-image specific neural network (DSNN) and (2) a feature-consistent generative adversarial network (FGAN). RNB: Residual Network Block.

disease (AD) and mild cognitive impairment (MCI) [1, 2]. The missing data problem is a common challenge in multi-modal neuroimaging studies, since subjects may lack a specific modality due to patient dropouts or poor data quality.

Conventional methods typically discard modality-incomplete subjects [1], which reduces the subjects that can be used to train a diagnosis model and hence may degrade the diagnostic performance. To utilize all available subjects, a cycle-consistent generative adversarial network (CGAN) has recently been employed to impute missing PET images [2, 3]. This model, however, equally treats all voxels in each 3D volume and thus ignores the *disease-image specificity* conveyed in multi-modal data. Specially, such disease-image specificity is two-fold, including (1) not all brain regions are associated with a specific disease, and (2) disease-related brain regions could be different in MRI and PET. Previous studies have shown that the diagnosis models can implicitly or explicitly represent the disease-image specificity [1]. Therefore, it is intuitively desirable to improve diagnostic performance by integrating diagnosis and image synthesis into a unified framework, i.e., imputing missing neuroimages in a diagnosis-oriented manner.

In this paper, we propose a *disease-image specific deep learning framework* for joint disease diagnosis and image synthesis using incomplete multi-modal neuroimages (see Fig. 1). As shown in Fig. 1, our method contains a disease-image specific neural network (DSNN) for diagnosis and a feature-consistent generative adversarial network (FGAN) for image synthesis. Herein, DSNN encodes disease-image specificity in its feature maps to assist the training of FGAN, while FGAN imputes missing images to improve the diagnostic performance of DSNN. Experimental results on 1,466 subjects suggest that our method can *not only* synthesize reasonable MR and PET images, *but also* achieve the state-of-the-art results in both tasks of AD identification and MCI conversion prediction.

2 Method

DSNN: It is considered that different regions in the brain vary in anatomy and/or function, and a specific brain disease is often associated with some specific regions. Hence, previous studies usually partition the brain into multiple regions-of-interest (ROIs) and then construct a disease classification model using the features extracted in pre-defined disease-associated ROIs [4,5]. By contrast, the proposed DSNN can directly extract features from the whole brain image and provide critical information implicitly (i.e., disease-related MRI/PET regions) to aid the image imputation conducted by FGAN.

We can decompose each MR/PET image X into (1) a disease-associated part and (2) a residual normal part. After feature extraction by $F(*)$, i.e., convolutional (Conv) layers in DSNN, the output feature map \mathbf{f} can be decomposed accordingly into the disease-associated part \mathbf{f}_d and the residual normal part \mathbf{f}_r , where $\mathbf{f} = F(X) = \alpha\mathbf{f}_d + (1 - \alpha)\mathbf{f}_r$ and α is a coefficient indicating the severity of disease-associated part for a respective subject. The diagnosis result should be independent of the residual normal part \mathbf{f}_r since \mathbf{f}_r is not related to the disease. Therefore, the response of the classifier $C(*)$ to the entire feature map is only associated with the disease-associated part, i.e., $C(\mathbf{f}) = C(\mathbf{f}_d)$. Note that, in DSNN, we propose a spatial cosine module to suppress the effects of α and \mathbf{f}_r , thus making the disease-related features conspicuous and easy to be captured.

Assume the feature map generated by the final Conv layer in DSNN to be $U = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K\}$, where each element is a vector corresponding to a spatial location in the brain. We first perform l_2 -normalization on each vector in U , and then concatenate them as the spatial representation of an MRI/PET scan:

$$\mathbf{u} = \left(\frac{\mathbf{v}_1^T}{\|\mathbf{v}_1\|_2}, \frac{\mathbf{v}_2^T}{\|\mathbf{v}_2\|_2}, \dots, \frac{\mathbf{v}_K^T}{\|\mathbf{v}_K\|_2} \right)^T. \quad (1)$$

This alleviates the influence from variation of α across different images. Theory to alleviate such a influence can refer to [6,7]. Suppose $C(*)$ is a classifier with a hyperplane parameter \mathbf{w} , defined as the cosine kernel:

$$C(\mathbf{u}; \mathbf{w}) = \frac{\mathbf{u}^T \mathbf{w}}{\|\mathbf{u}\|_2 \|\mathbf{w}\|_2} = \frac{\mathbf{u}^T}{\|\mathbf{u}\|_2} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|_2}, \quad (2)$$

which is equivalent to the product of l_2 -normalized \mathbf{u} and \mathbf{w} , both having the constant unit norm. The constant norm forces $C(*)$ to focus on the disease-associated part since all features have the same norm after l_2 -normalization (in Eq. 1), and thus suppresses the influence of the residual normal part.

As shown in the bottom left part of Fig. 1, our DSNN contains sequentially (1) a feature extraction module and (2) a spatial cosine module. The first module has 5 Conv layers, with 16, 32, 64, 64, and 64 channels, respectively. The first 4 and the last Conv layers are respectively followed by the max-pooling and average-pooling with the stride of 2 and the size of 3×3 . By inputting a neuroimage, the feature extraction module outputs feature maps at each Conv layer. The spatial cosine module first l_2 -normalizes the feature vectors in the feature map

of the 5th Conv layer and concatenates them to a spatial representation, and then utilizes a fully-connected layer with cosine kernel (as Eq. 2) to compute the probability score of a subject belonging to a particular category.

FGAN: Motivated by the fact that a pair of MR and PET images (scanned from the same subject) have underlying relevance but probably different disease-associated regions, we accordingly develop a feature-consistent generative adversarial network (FGAN) to synthesize a missing PET image based on its corresponding MRI (see the top left in Fig. 1). Especially, FGAN contains a feature-consistent component to encourage the feature maps of a synthetic image produced by the diagnosis model (i.e., DSN containing disease-image specificity) to be consistent with the feature maps of its corresponding real image.

Denote X_M and X_P as the MRI and PET scans from the same subject. We aim to learn an image generator $G_M : X_M \rightarrow X_P$ to impute the missing PET image via $G_M(X_M)$. An inverse mapping $G_P : X_P \rightarrow X_M$ ($G_P = G_M^{-1}$) can be also learned to build a bidirectional mapping between MRI and PET [2]. Basically, a discriminator is needed to encourage that the distribution of synthetic scans (e.g., $G_M(X_M)$) is indistinguishable from the distribution of its corresponding real images (e.g., X_P). We further introduce a *feature-consistent component* to enforce that the disease-image specificity is consistent between the synthetic scan (e.g., $G_M(X_M)$) and its real scan (e.g., X_P). Accordingly, two kinds of loss are incorporated into our FGAN, including an adversarial loss and a feature-consistent loss. Specifically, the *adversarial loss* is defined as

$$\begin{aligned} \mathcal{L}_g(X_M, X_P; G_M, G_P, D_M, D_P) = & (\log(D_P(X_P)) + \log(1 - D_P(G_M(X_M)))) \\ & + (\log(D_M(X_M)) + \log(1 - D_M(G_P(X_P))))), \end{aligned} \quad (3)$$

where D_M and D_P are two discriminators that can identify whether an input image is real or synthetic. The proposed *feature-consistent loss* is defined as

$$\begin{aligned} \mathcal{L}_c(X_M, X_P; G_M, G_P, F_M, F_P) = & \|F_M(G_P(X_P)) - F_M(X_M)\| \\ & + \|F_P(G_M(X_M)) - F_P(X_P)\|, \end{aligned} \quad (4)$$

where F_M and F_P are the proposed feature-consistent components to encourage that a pair of synthetic and real scans from the same modality share the same disease-image specificity. Finally, the overall loss function of FGAN is defined as

$$\begin{aligned} \mathcal{L}(X_M, X_P; G_M, G_P, D_M, D_P, F_M, F_P) = & \mathcal{L}_c(X_M, X_P; G_M, G_P, F_M, F_P) \\ & + \mathcal{L}_g(X_M, X_P; G_M, G_P, D_M, D_P). \end{aligned} \quad (5)$$

As shown in the top left part of Fig. 1, our FGAN contains three components, including (1) two generators (i.e., G_M and G_P), (2) two adversarial discriminators (i.e., D_M and D_P), and (3) two feature-consistent components (i.e., F_M and F_P). Specifically, each generator (e.g., G_M) consists of three Conv layers (with 8, 16, and 32 channels, respectively) to extract the knowledge of images in the original domain (e.g., X_M), six residual network blocks (RNBs) [8] to transfer the knowledge from the original domain (e.g., X_M) to the target domain (e.g., X_P), and two deconvolutional (Deconv) layers (with 32 and 16 channels, respectively)

and one Conv layer (with 1 channel) to construct the image in the target domain (e.g., X_P). Each discriminator (e.g., D_P) contains five Conv layers, with 16, 32, 64, 128, and 1 channel(s), respectively. It outputs an indicator to tell whether the input pair of real image (e.g., X_P) and synthetic image (e.g., $G_M(X_M)$) are distinguishable (output: 0) or not (output: 1). Each feature-consistent component (e.g., F_P) contains two subnetworks with shared architecture and parameters from the feature extraction module in our DSNN. It inputs a pair of real image (e.g., X_P) and synthetic image (e.g., $G_M(X_M)$), and then outputs a differential score to indicate the similarity between the feature maps of the real and its corresponding synthetic image. In this way, the disease-image specificity learned in DSNN can be used to aid the image imputation process in FGAN, by which the modality-specific disease-related brain regions could be more effectively synthesized. In turn, the synthetic images could be more relevant to brain disease diagnosis, thus effectively improving the performance of DSNN.

Implementation: We first train F_M and F_P via DSNN for 40 epochs using complete subjects (i.e., those with paired PET and MR images). We then train D_M and D_P by minimizing $-\mathcal{L}_g(*)$ with fixed G_M and G_P , and train G_M and G_P by minimizing $\mathcal{L}(*)$ with fixed D_M and D_P , iteratively, for 100 epochs. After that, we retrain DSNN for 40 epochs using both real and synthetic data. The Adam solver [9] is used with a batch size of 1 and a learning rate of 2×10^{-3} .

3 Experiments

Materials and Experimental Setup: We evaluated the proposed method on two subsets of ADNI [10], including ADNI-1 and ADNI-2. Subjects were divided into four categories: (1) AD, (2) cognitively normal (CN), (3) progressive MCI (pMCI) that would progress to AD within 36 months after baseline, and (4) static MCI (sMCI) that would not progress to AD. After removing subjects that exist in both ADNI1 and ADNI2 from ADNI2, there are 205 AD, 231 CN, 165 pMCI and 147 sMCI subjects in ADNI-1, while there are 162 AD, 209 CN, 89 pMCI and 258 sMCI subjects in ADNI-2. All subjects in ADNI-1 and ADNI-2 have baseline MRI data, while only 403 subjects in ADNI-1 and 595 subjects in ADNI-2 have PET images. For image pre-processing, we performed skull-stripping, intensity correction and spatial normalization for all images. Each PET image was aligned to its corresponding MRI scan. Hence, there is spatial correspondence between a pair of MRI and PET scans for each subject.

We performed two groups of experiments in this work. In the *first* group, we evaluated the quality of synthetic images generated by our method. In the *second* group, we compared our method with state-of-the-art methods in both tasks of AD identification (AD vs. CN classification) and MCI conversion prediction (pMCI vs. sMCI classification). Subjects from ADNI-1 were used for training the models, while those from ADNI-2 for the independent test. The first group of experiments were carried on only real MRI and PET scans, while the second group of experiments used both real and synthetic multi-modal neuroimages.

Evaluation of Synthetic Neuroimages: We first evaluated the quality of synthetic images generated by our **FGAN** method. Two generative models were compared with FGAN, including (1) a conventional **GAN** with only the adversarial loss, and (2) the cycle-consistent GAN (**CGAN**) [2]. We trained these three models (i.e., GAN, CGAN, and FGAN) using the subjects with MRI and PET scans in ADNI-1, and tested the trained models on complete subjects (with both MRI and PET) in ADNI-2. We used three metrics to measure the quality of synthetic images generated by different methods, including (1) the mean absolute error (MAE), (2) peak signal-to-noise ratio (PSNR), and (3) structural similarity index measure (SSIM) [11]. To evaluate the reliability of synthetic MR and PET images in disease diagnosis, we further reported the values of the area under receiver operating characteristic (AUC) achieved by our DSNN on both classification tasks of AD vs. CN (AUC*) and pMCI vs. sMCI (AUC[†]). For a fair comparison, we first trained two DSNN models on ADNI-1 using real MRI and real PET scans, respectively. Then, we applied these DSNNs to subjects in ADNI-2 represented by synthetic MR and PET images, respectively.

Table 1. Results (% except PSNR) of image synthesis achieved by three different methods for MRI and PET scans of subjects in ADNI-2, with the models trained on ADNI-1.

Method	Synthetic MRI					Synthetic PET				
	MAE	SSIM	PSNR	AUC*	AUC [†]	MAE	SSIM	PSNR	AUC*	AUC [†]
GAN	15.03	55.56	23.48	66.83	53.64	10.54	57.70	27.15	60.47	51.92
CGAN	14.09	59.38	23.96	71.00	57.20	9.72	61.46	28.16	65.91	58.43
FGAN (Ours)	12.61	64.04	25.10	92.98	77.61	8.03	68.17	29.62	83.80	71.27

Results of image synthesis are reported in Table 1. Three interesting observations can be found from Table 1. *First*, FGAN and CGAN generally yield better results than GAN, suggesting that the cycle-consistent loss and feature-consistent loss help synthesize images with higher quality. *Second*, our FGAN consistently outperforms CGAN in synthesizing both MRI and PET scans in terms of three metrics (i.e., MAE, SSIM and PSNR). It reveals that the proposed feature-consistent loss in FGAN is more effective in improving the image quality of synthetic images, compared with the cycle-consistent loss used in CGAN. *Besides*, AUC values obtained by using FGAN-based synthetic images are significantly better than those using CGAN-based and GAN-based synthetic images. This implies that our FGAN can generate diagnosis-oriented images, thus helping boost the performance of brain disease diagnosis. In Fig. 2, we further visualize the synthetic images generated by three methods and their corresponding real/ground-truth images for a typical subject in ADNI-2. Figure 2 suggests that the images synthesized by our FGAN are more visually similar to the ground truth (especially for the hippocampus regions, see yellow and red squares in Fig. 2), compared with those generated by GAN and CGAN.

Evaluation of Disease Identification: We further evaluated our DSNN method in both tasks of AD identification (AD vs. CN) and MCI conversion prediction (pMCI vs. sMCI). The proposed DSNN is first compared with two conventional methods using concatenated MRI and PET features, i.e., (1) ROI method [12], and (2) patch-based morphology (PBM) [13]. We also compared DSNN with two deep learning methods, i.e., (3) landmark-based deep multi-instance learning (LDMIL) [5], and (4) a variant of DSNN (called DSNN1) that globally averages the feature map of the fifth Conv layer and uses a fully connected layer for classification (instead of using the spatial cosine module in DSNN). In DSNN/DSNN1, a pair of MR and PET images are fed into two parallel DSNN/DSNN1 models to generate two probability scores, followed by averaging these scores for classification. These five methods utilize all subjects with both real multi-modal (i.e., MRI and PET) scans and synthetic PET images generated by our FGAN. We also performed experiments on complete subjects (with only real MRI and PET scans), and denoted the corresponding methods as “-C”. We employ six metrics for performance evaluation in disease diagnosis, including AUC, accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-Score (F1S), and Matthews correlation coefficient (MCC) [14]. Disease classification results were reported in Table 2.

From Table 2, we can see that the overall performance of our DSNN method is superior to four competing methods using both real and synthetic data in terms of six metrics. For instance, our DSNN method achieves the best AUC value (83.94%) in pMCI vs. sMCI classification. This implies that DSNN is reliable in predicting the progression of MCI patients, which is potentially very useful in practice. *Besides*, in both classification tasks, three deep learning methods (i.e., LDMIL, DSNN1, and DSNN) generally outperform two conventional approaches (i.e., ROI and PBM) that use hand-crafted features. This suggests that integrating feature extraction and classifier construction into a unified framework (as we do in DSNN) can boost the performance of brain disease diagnosis. *Furthermore*, DSNN yields better diagnostic performance than both LDMIL and its variant (i.e., DSNN1 with simple average pooling). This implies that the proposed spatial cosine kernel provides a more efficient strategy to capture the disease-image specificity embedded in neuroimages, compared with the cases of using pre-defined disease-related regions in LDMIL and without mining the disease-image specificity in DSNN1. *In addition*, the methods using both real and synthetic images consistently outperform their counterparts that utilize just complete subjects with real MRI and PET scans, suggesting that the neuroimages generated by our FGAN are useful in promoting the performance of brain disease diagnosis. Note that we didn’t use synthetic MRIs generated by FGAN in Table 2, because all subjects in ADNI have real MRIs. Using both real and our synthesized MRIs for training, the proposed DSNN method achieved an AUC value of 84.24% in pMCI vs. sMCI classification, which is slightly better than that (83.94%) of DSNN without using synthetic MRIs generated by FGAN.

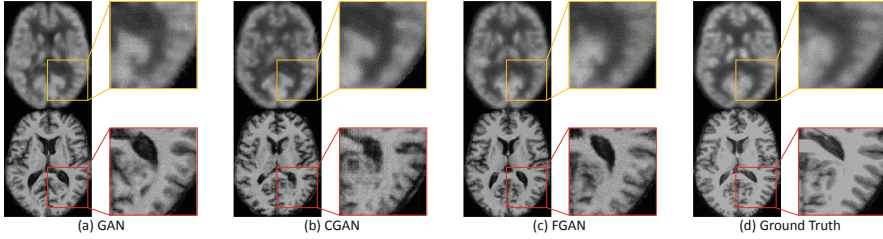


Fig. 2. PET and MRI scans synthesized by three methods for a typical subject (Roster ID: 4029) in ADNI-2, along with their corresponding ground-truth images.

Table 2. Diagnosis results (%) achieved by five different methods using complete subjects (denoted as “-C” with real MRI and PET scans) and all subjects (with both real images and synthetic PET images generated by FGAN) in two classification tasks.

Method	AD vs. CN classification						pMCI vs. sMCI classification					
	AUC	ACC	SPE	SEN	F1S	MCC	AUC	ACC	SPE	SEN	F1S	MCC
ROI-C	88.95	81.49	82.99	80.32	79.74	62.92	69.17	66.92	42.50	77.78	44.16	20.74
PBM-C	87.87	82.39	80.27	84.04	80.00	64.27	68.04	68.85	60.00	72.78	54.24	31.28
LDMIL-C	95.64	90.75	88.44	92.55	89.35	81.18	82.62	76.92	71.25	79.44	65.52	48.67
DSNN1-C	94.40	89.25	88.44	89.89	87.84	78.22	78.70	75.00	67.50,	78.33	62.43	44.13
DSNN-C (Ours)	95.78	90.45	89.80	90.96	89.19	80.64	82.74	76.15	76.25	76.11	66.30	49.33
ROI	90.51	83.42	84.24	82.78	81.76	66.69	72.31	71.30	52.81	77.73	48.70	29.12
PBM	91.46	84.49	81.21	87.08	82.21	68.48	72.79	71.88	64.04	74.61	54.03	35.37
LDMIL	96.76	91.71	88.48	94.26	90.40	83.17	83.89	79.71	71.91	82.42	64.65	51.13
DSNN1	94.72	89.57	85.45	92.82	87.85	78.82	79.98	76.52	69.66	78.91	60.49	44.98
DSNN (Ours)	97.02	92.25	90.91	93.30	91.18	84.27	83.94	80.00	70.79	83.20	64.61	51.20

4 Conclusion

We proposed a disease-image specific deep learning framework for joint disease diagnosis and image synthesis with incomplete multi-modal neuroimaging data, where a diagnosis network is employed to provide disease-image specificity to an image synthesis network. Specifically, we designed a disease-image specific neural network (DSNN) and trained it on the entire image to implicitly capture the disease-associated information conveyed in MRI and PET. We further developed a feature-consistent generative adversarial network (FGAN) to synthesize missing neuroimages, by encouraging DSNN feature maps of each synthetic image and the respective real image to be consistent. Experiments on ADNI demonstrate that our method generates reasonable neuroimages and also achieves the state-of-the-art performance in AD identification and MCI conversion prediction.

References

1. Zhou, J., Yuan, L., Liu, J., Ye, J.: A multi-task learning formulation for predicting disease progression. In: ACM SIGKDD, pp. 814–822. ACM (2011)
2. 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. *MICCA I*, 455–463 (2018)
3. Qiao, T., Zhang, J., Xu, D., Tao, D.: MirrorGAN: learning text-to-image generation by redescription. arXiv preprint [arXiv:1903.05854](https://arxiv.org/abs/1903.05854) (2019)
4. Wachinger, C., Salat, D.H., Weiner, M., Reuter, M.: Whole-brain analysis reveals increased neuroanatomical asymmetries in dementia for hippocampus and amygdala. *Brain* **139**(12), 3253–3266 (2016)
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. Pan, Y., Xia, Y., Shen, D.: Foreground fisher vector: encoding class-relevant foreground to improve image classification. *IEEE Trans. Image Process.* (2019)
7. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the fisher vector: theory and practice. *Int. J. Comput. Vis.* **105**(3), 222–245 (2013)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
10. Jack, C., Bernstein, M., Fox, N., et al.: The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imag.* **27**(4), 685–691 (2008)
11. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: ICPR, pp. 2366–2369 (2010)
12. Rusinek, H., et al.: Alzheimer disease: measuring loss of cerebral gray matter with MR imaging. *Radiology* **178**(1), 109–114 (1991)
13. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Collins, D.L.: Simultaneous segmentation and grading of anatomical structures for patient’s classification: application to Alzheimer’s disease. *NeuroImage* **59**(4), 3736–3747 (2012)
14. Koyejo, O.O., Natarajan, N., Ravikumar, P.K., Dhillon, I.S.: Consistent binary classification with generalized performance metrics. In: NIPS, pp. 2744–2752 (2014)