

ENHANCING ALZHEIMER’S DISEASE DIAGNOSIS VIA HIERARCHICAL 3D-FCN WITH MULTI-MODAL FEATURES

Chao Liu¹, Xiaodong Yang¹, Dading Chong¹, Wenwu Wang², and Liang Li^{1,*}

¹School of Psychological and Cognitive Sciences, Peking University

²Centre for Vision, Speech and Signal Processing, University of Surrey

ABSTRACT

Alzheimer’s disease (AD) is an incurable, progressive neurological disorder of the human brain related to loss of memory, commonly seen in the elderly population. Accurate detection of AD can help with proper treatment and prevent brain function damage. Existing CNN-based methods need to predetermine informative locations in sMRI, which means the stage of distinguishing lesions is separated from the later stages of feature extraction and classifier construction. In this paper, a novel “two-stage” framework based on a hierarchical 3D fully convolutional network (H-3D-FCN) is proposed to automatically identify discriminative local patches and regions in the sMRI. We further optimize the diagnosis performance by constructing a multi-layer perceptron (MLP) model which combines the multi-modal features (e.g., MMSE score, age, gender, APOE 4) with the risk probability maps (RPMs) generated from the H-3D-FCN model. Experiments on three typical AD datasets, namely, ADNI, AIBL, and NACC, show that our method achieves state-of-the-art performance as compared with recent baselines.

Index Terms— Alzheimer’s disease diagnosis, H-3D-FCN, Multimodality learning, Multi-layer perceptron

1. INTRODUCTION

Alzheimer’s disease (AD) is a severe form of mental dementia. Around 640 million people worldwide continue to suffer from AD and develop dementia symptoms that worsen over time. Therefore, accurate diagnosis of AD is important since treatment may be most effective if introduced as early as possible [1]. In practice, a standardized diagnostic process relies on skilled clinicians, based on clinical history, psychometric assessment such as Mini-Mental State Examination (MMSE), and brain imaging such as structural magnetic resonance imaging (sMRI). However, clinicopathological studies have shown that clinicians’ diagnostic sensitivity ranged from 70.9% to 87.3%, and specificity ranged from 44.3% to 70.8% [2], the wider range of sensitivity and specificity means that clinicians have a relatively high risk of leading misdiagnosis and missed diagnosis [3].

Recently, AD diagnosis has been successfully applied to the convolutional neural network (CNN) and shows promising results [4, 5]. Preliminary studies, however, indicate that, due to subtle structural changes in the brain, these approaches are not successful in detecting early-stage AD [6]. To address this challenge, either informative regions (e.g., hippocampus [7]) or patches (e.g., positioned by an anatomical landmark detector [8]) are determined using domain knowledge and expert experience before applying CNN-based methods. That is, the differential localization stage of brain lesions [9] is methodologically independent of the later stages of feature extraction and classifier construction, which may hinder the effectiveness of CNN in the diagnosis of AD [10].

In this study, we present a novel deep learning framework to unify discriminant lesion location with feature extraction and classifier construction to improve the performance for sMRI-based AD diagnosis. Specifically, a hierarchical 3D fully convolutional network (H-3D-FCN) is proposed to enhance the feature representation of AD by automatically identifying the differentiated locations in sMRI of the whole brain. The framework of H-3D-FCN includes three modules, namely, feature encoder, decoder, and classifier. In addition, a “two-stage” training scheme is adopted to optimize the diagnosis performance by using multi-modal diagnosis features. In stage 1, we use randomly-sampled sub-volume brain images as input to train the H-3D-FCN model. Then, the full-sized brain images are fed into the training model, and the whole-brain risk probability maps (RPMs) are generated by softmax to infer local patterns of the cerebral structure reflecting an overall disease state.

Inspired by a reference indicator used by clinicians to diagnose AD, the RPMs generated above are integrated with other modal features through a multi-layer perceptron (MLP) in stage 2, such as specific psychometric assessment (MMSE score [11]), demographic information (age, gender) and genetic information for the therapeutic target of AD (APOE 4 [12]). Following training and internal testing on ADNI dataset [13], we further validated the performances on AIBL [14], and NACC [15]. Our method is shown to give an accuracy of 94.1%, 95.3%, and 87.4%, respectively, on ADNI, AIBL, and NACC, significantly better than three recent state-of-the-art methods.

This study was supported by the National Natural Science Foundation of China (31771252) to Liang Li. (*Corresponding author: Liang Li.)

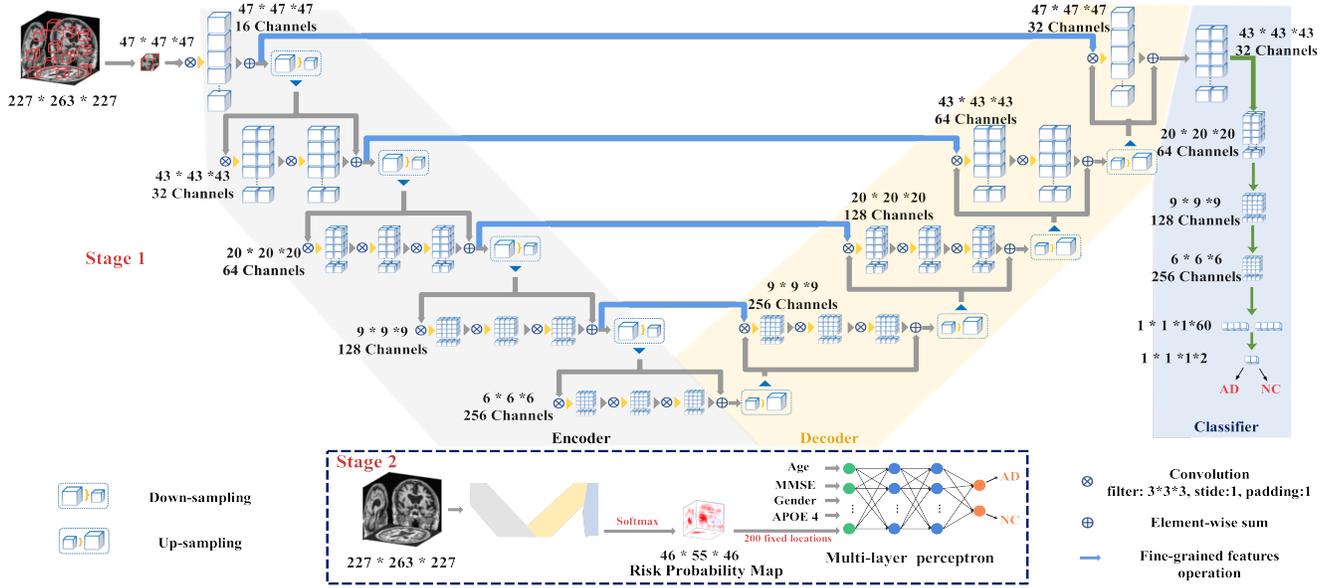


Fig. 1. The proposed “two-stage” AD diagnostic framework is composed of three modules, namely, a feature encoder, a decoder and a classifier. In stage 1, the AD status corresponding to the individual subject is used as the output of the classification model. In stage 2, subject-specific brain risk probability maps (RPMs) are generated, then 200 fixed locations are selected from RPMs and passed to the multi-layer perceptron with multi-modal information for further binary classification of the disease status.

2. PROPOSED METHOD

Our proposed hierarchical 3D-FCN framework is shown in Fig. 1, which includes three modules, i.e., an Encoder, a Decoder, and a Classifier. Meanwhile, referring to the reference indicators of clinicians for AD diagnosis, the “two-stage” training scheme is adopted to boost the diagnosis performance by utilizing multi-modal diagnosis features.

2.1. Network architecture

Encoder The gray shadow in Fig. 1 shows the encoder, consisting of five successive stages of convolution and down-sampling that operate at different resolution of 3D feature. Each stage, shown in Fig. 2, consists of two sub-modules, namely, the residual module (RM) and down-sampling module. The RM, inspired by He et al. [16], is designed using a residual function to obtain better features without increasing the receptive field, as shown in the bottom dashed box in Fig. 2, where the features from the upper layer are processed by a series of convolutional layers with a $3 \times 3 \times 3$ kernel applied with stride 1 and padding 1. The processed convolutional features are fused with the original input by element-wise operation, which result in outputs of equal size as the inputs. After that, the fused features are down-sampled using the 3D max-pooling operation, defined as F_{ds} , which reduces the size of the input and increases the receptive field of the features calculated in the subsequent network layer. In detail, we define the k -th feature map of the $(l-1)$ -th stage as F_k^{l-1} , which is connected with the j -th feature map of the l -th stage. To

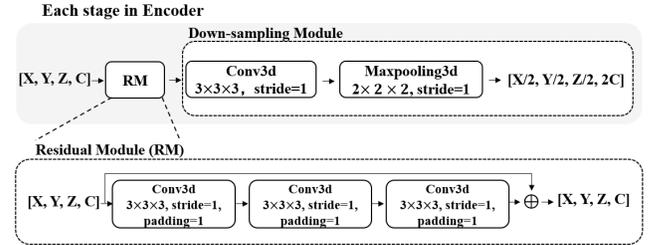


Fig. 2. Each stage in Encoder contains two sub-modules: residual module (RM) and down-sampling module. “Conv3d” stands for Conv3d-BN-PReLU, $[X, Y, Z, C]$ represents the size of 3D feature and its channels.

simplify the formula, we consider one convolutional layer in the residual module before applying the down-sampling module. Therefore, the procedure of each stage of the encoder is defined as:

$$F_j^l = PReLU \left(Conv3d_{3 \times 3 \times 3} (F_k^{l-1}) \right) \oplus F_k^{l-1} \quad (1)$$

$$F_{ds} = PReLU \left(Maxpooling_{2 \times 2 \times 2} (F_j^l) \right) \quad (2)$$

Decoder The yellow shadow in Fig. 1 shows the decoder, consisting of four stages, which aim to extract features, and expand the spatial support of lower resolution feature maps, and to obtain semantic information of features. Each stage, depicted in Fig. 3, consists of three sub-modules, namely, up-sampling module, the fine-grained module (FGM), and the residual module (RM), same as in the encoder. Specifically,

a 3D dilated convolution operation is employed in the up-sampling module to increase the size of inputs and support the exponential expansion of the receptive field without loss of resolution or coverage. The parameters of each stage in this module are shown in the table of Fig. 3. The FGM (shown in blue arrow in Fig. 1) aims to improve the quality of the feature representation, indicated in the dashed frame in the right corner of Fig. 3. More specifically, a concatenation operation is utilized to integrate the shallower layers with detailed features in the encoder and the deeper layers with abstract features in the decoder. Afterwards, the concatenated features are input into a 3D convolutional layer to retain the same channels with next stage, and then, as input to the residual module. In detail, we define the i -th feature map of the $(l-1)$ -th stage as F_i^{l-1} , which is connected with the h -th feature map of the l -th stage. To simplify the formula, we also consider one convolutional layer in RM before applying the up-sampling module. Each stage of the decoder is defined as:

$$F_{concat} = Conv3d_{1 \times 1 \times 1} \left(Concat(F_h^{l-1}, F_j^l) \right) \quad (3)$$

$$F_h^l = PReLU(Conv3d_{3 \times 3 \times 3}(F_{concat})) \oplus F_h^{l-1} \quad (4)$$

$$F_{upsample} = PReLU(Dilated_{3 \times 3 \times 3}(F_h^l)) \quad (5)$$

Classifier The blue shadow in the right of Fig. 1, shows the classifier, consisting of six convolutional blocks. Each of the first four convolutional blocks contains a 3D convolutional layer with a $3 \times 3 \times 3$ kernel applied with stride 1, appended with 3D max-pooling (kernel = 2, stride = 1, padding = 0), 3D batch-normalization, PReLU, and Dropout ($P = 0.5$). Different from the first four convolutional blocks, the kernel sizes have been changed to 6 and 1 in the last two convolutional blocks respectively. Finally, two channels' features are generated by the classifier, a softmax function is adopted to predict the specific mental status (AD/NC).

The framework uses a “two-stage” training scheme to incorporate multi-modal features to improve diagnostic performance. Considering the small scale of the sMRI dataset of AD, we employed a patch-wise data augmentation strategy in stage 1, in which 3000 samples (sub-volume is $47 \times 47 \times 47$ voxel) were randomly selected from each whole-brain sMRI as input, and trained the binary classification model by generating two channels' features with volume $1 \times 1 \times 1$. In stage 2, the full-sized sMRI ($227 \times 263 \times 227$) are then fed into the training model in stage 1 to yield a 3D tensor of size $46 \times 55 \times 46$ with two channels, which can be translated to risk probability maps (RPMs) by a softmax function. As shown at the bottom of Fig. 1, RPMs can be used to infer the local pattern of brain structure and estimate the overall disease state. Then, the whole-brain RPMs of sMRI were generated for all subjects using the model we trained in stage 1. In addition, as a reference indicator used by clinicians for AD diagnosis, demographic information and clinical characteristic of the subjects are related to the state of illness. Therefore, using these multi-modal features can further improve the classification performance.

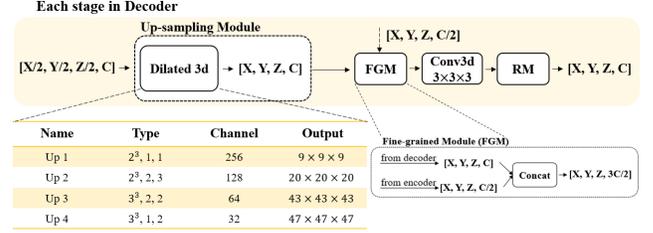


Fig. 3. Each stage in Decoder contains three sub-modules respectively, up-sampling module, fine-grained module (FGM), and RM. “Dilated 3d” stands for Dilated 3d-BN-PReLU, $[X, Y, Z, C]$ represents the size of 3D feature and its channels, the specific parameters of each stage of “Dilated 3d” are also shown in the table, where the parameter “Type” represents kernel size, stride, and dilation.

2.2. MLP model

As shown in Fig. 1, we use a multi-layer perceptron (MLP) to fuse the multi-modal features, including an input layer, two hidden layers and an output layer. Among them, principal component analysis (PCA) is employed in the hidden layers to decompose the feature space. The activation function and loss function used are ReLU and softmax cross-entropy with logits, respectively. More precisely, we selected RPM voxels from 200 fixed locations which have higher Matthew’s correlation coefficient values, and use the feature vectors extracted from these locations, combined with other modal features, as input to the MLP for classification. Specifically, the feature vectors from RPM, MMSE, age, gender, and APOE 4 are used as inputs of the MLP for binary classification of the final mental status of the subject.

3. EXPERIMENTS AND RESULTS

This section presents performance evaluations of our proposed method and comparisons with existing methods.

3.1. Datasets and Subject Selection

Three typical AD datasets used in our experiments are, **ADNI** [13], **AIBL** [14], **NACC** [15]. Some subjects are selected from these three datasets in terms of the following criteria. Firstly, age should be 55 or above. Secondly, 1.5 T with T1-weighted MRI scans taken with ± 6 months need to be diagnosed with AD or normal condition (NC). Thirdly, subjects with mixed dementia, non-AD dementia, history of severe head trauma, severe depression, history of stroke and brain tumor, and major systemic diseases are excluded. According to these criteria, 415, 381, and 543 subjects were selected from ADNI, AIBL, and NACC, respectively. Meanwhile, the specific information of MMSE score, gender, age and APOE 4 of each subject was obtained from these datasets.

Table 1. The performance of our method on ADNI test, AIBL, NACC (%)

Dataset	Model	ACC	SEN	SPEC
ADNI test	*sMRI	85.2±2.2	82.1±2.4	89.2±1.5
AIBL	*sMRI	90.4±1.4	70.3±2.0	94.2±1.2
NACC	*sMRI	85.7±3.1	81.3±2.6	88.4±3.3
ADNI test	+MLP	94.1±1.9	96.2±2.3	96.5±2.4
AIBL	+MLP	95.3±2.8	89.8±1.6	95.3±2.3
NACC	+MLP	87.4±4.0	93.7±1.5	89.5±4.4

Table 2. The comparison of performances on NACC (%)

Method	ACC	SEN	SPEC
Qiu et al. [17]	84.3±4.7	90.1±2.4	80.5±4.3
Feng et al. [5]	80.3±3.6	82.3±5.9	86.3±5.9
Korolev et al. [18]	76.4±4.2	83.3±3.7	82.4±3.5
Ours: *sMRI	85.7±3.1	81.3±2.6	88.4±3.3
Ours: +MLP	87.4±4.0	93.7±1.5	89.5±4.4

3.2. Data Preprocessing

The sMRI data from the three datasets are all in NIFTI format. We preprocessed the data with alignment, registration, and intensity normalization, respectively. Firstly, we conduct Anterior Commissure (AC)-Posterior Commissure (PC) alignment and re-sample the data to $227 \times 263 \times 227$ via the MIPAV Software [19]. Then, the HAMMER algorithm [20] is utilized to spatially register the sMRI data with respect to the MNI152 template using the FSL package [21]. Finally, we normalize tissue intensities inhomogeneity using the N3 algorithm [22].

3.3. Evaluation Metric

For evaluation, we use three performance metrics, namely, accuracy (ACC), sensitivity (SEN), and specificity (SPEC) [17]. We perform five independent tests, and present the mean and standard deviation over these tests.

3.4. Implementation Details

Following training and internal testing on ADNI which randomly split it in the ratio of 3:1:1 for training, validation, and testing, and we further validated the performances on AIBL, and NACC. Specifically, in stage 1, the training set of ADNI was used to train our proposed method for 3,000 epochs. We choose Adam optimizer for optimization and categorical cross entropy as the loss function [23]. Meanwhile, batch size was set to 40 and base learning rate was set as 0.0001 for the first 1000 epochs and then fixed at 0.001 for the rest. In stage 2, a gradient descent algorithm [24] is applied to train the MLP model and identify the weights and biases of each layer during the back-propagation process.

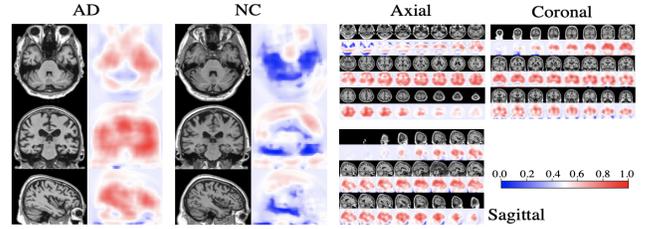


Fig. 4. The RPMs of two specific subjects with AD and NC from ADNI (left), and the RPM of this AD patient on axial, coronal, and sagittal planes (right).

3.5. Experimental Results and Analysis

Table 1 shows the results of our method on the three selected datasets. Specifically, two models are tested, respectively, the model of merely sMRI as input (*sMRI) and the model appended with MLP (+MLP). The average accuracy in terms of *sMRI on ADNI test, AIBL, and NACC are 85.2%, 90.4%, and 85.7%, respectively. For the +MLP model, the average accuracy on ADNI test, AIBL, and NACC are 97.1%, 95.3%, and 87.4%, respectively. It can be seen that the +MLP model is much better than the *sMRI model, illustrating that the fusion of multi-modal clinical features can further improve the performance of AD diagnosis. Meanwhile, the sensitivity and specificity of ours based on the +MLP model is clearly higher and has smaller range fluctuation than clinicians [2].

Furthermore, for making a fair comparison, we unified the criterion for data selection, then we compared the performances of previous promising methods with ours on these three datasets. Our experimental results have shown that our method is much better than these baseline methods among three datasets. We list the specific results on a more challenging dataset, NACC, as shown in Table 2, manifests our method performs more effectively.

Finally, we visualize the RPMs generated by our method for two specific subjects with AD and NC from the ADNI dataset, shown in the left of Fig. 4. Some potential sick brain areas can be seen clearly from the RPMs of AD in Fig. 4 such as temporal lobes, hippocampus, cingulate cortex. We also visualize the RPMs of this AD subject on axial, coronal, and sagittal planes on the right of the Fig. 4, showing the distribution of the potential risk of illness.

4. CONCLUSION

In this paper, ‘‘H-3D-FCN’’ has been proposed for automatic identification of discriminative local patches and regions in the whole-brain sMRI for AD diagnosis. Meanwhile, the ‘‘two-stage’’ training scheme combined with multi-modal features is adopted to improve the diagnosis performance by employing the multi-layer perceptron (MLP). Experimental results validate the effectiveness of the proposed method, which offers promising results on ADNI, AIBL, and NACC datasets.

5. REFERENCES

- [1] Jason Weller and Andrew Budson, "Current understanding of alzheimer's disease diagnosis and treatment," *F1000Research*, vol. 7, 2018.
- [2] Thomas G Beach, Monsell, et al., "Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010," *Journal of Neuropathology and Experimental Neurology*, vol. 71, no. 4, pp. 266–273, 2012.
- [3] Erin P Balogh, Bryan T Miller, and John R Ball, "Improving diagnosis in health care," 2015.
- [4] Maheshappa Kruthika, Rajeswari et al., "Cbir system using capsule networks and 3d cnn for alzheimer's disease diagnosis," *Informatics in Medicine Unlocked*, vol. 14, pp. 59–68, 2019.
- [5] Chiyu Feng, Elazab, et al., "Deep learning framework for alzheimer's disease diagnosis via 3d-cnn and fsbi-lstm," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.
- [6] Fennema-Notestine et al., "Structural mri biomarkers for preclinical and mild alzheimer's disease," *Human Brain Mapping*, vol. 30, no. 10, pp. 3238–3253, 2009.
- [7] Alexander Khvostikov, Aderghal, et al., "3d cnn-based classification using smri and md-dti images for alzheimer disease studies," *arXiv preprint arXiv:1801.05968*, 2018.
- [8] Mingxia Liu, Jun Zhang, Ehsan Adeli, and Dinggang Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Medical Image Analysis*, vol. 43, pp. 157–168, 2018.
- [9] Bolei Zhou, Aditya Khosla, et al., "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929.
- [10] Chunfeng Lian et al., "Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 880–893, 2020.
- [11] Marshal F Folstein, Lee N Robins, and John E Helzer, "The mini-mental state examination," *Archives of General Psychiatry*, vol. 40, no. 7, pp. 812–812, 1983.
- [12] David A Sanan et al., "Apolipoprotein e associates with beta amyloid peptide of alzheimer's disease to form novel monofibrils. isoform apoe4 associates more efficiently than apoe3.," *Journal of Clinical Investigation*, vol. 94, no. 2, pp. 860–869, 1994.
- [13] Clifford R Jack Jr, Bernstein, et al., "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.
- [14] Kathryn A Ellis, Bush, et al., "The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009.
- [15] Duane L Beekly, Ramos, et al., "The national alzheimer's coordinating center (nacc) database: the uniform data set," *Alzheimer Disease & Associated Disorders*, vol. 21, no. 3, pp. 249–258, 2007.
- [16] Kaiming He, Zhang, et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [17] Shangran Qiu, Prajakta S Joshi, et al., "Development and validation of an interpretable deep learning framework for alzheimer's disease classification," *Brain*, 2020.
- [18] Sergey Korolev, Amir Safiullin, Belyaev, et al., "Residual and plain convolutional neural networks for 3d brain mri classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 835–838.
- [19] Pierre-Louis Bazin, Cuzzocreo, et al., "Volumetric neuroimage analysis extensions for the mipav software package," *Journal of Neuroscience Methods*, vol. 165, no. 1, pp. 111–121, 2007.
- [20] Dinggang Shen and Christos Davatzikos, "Hammer: hierarchical attribute matching mechanism for elastic registration," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.
- [21] Yongyue Zhang, Brady, et al., "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [22] John G Sled, Alex P Zijdenbos, and Alan C Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.
- [23] Zhilu Zhang and Mert Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Advances in Neural Information Processing Systems*, 2018, pp. 8778–8788.
- [24] Sebastian Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.