Multi-modal Image Prediction via Spatial Hybrid U-Net

Akib Zaman[1(✉)], Lu Zhang[1], Jingwen Yan[2], and Dajiang Zhu[1]

[1] The University of Texas at Arlington, Arlington, USA
akbzmn@gmail.com
[2] Indiana University-Purdue University Indianapolis, Indianapolis, USA

Abstract. Cortical folding patterns and white matter connectivity together compose the structural organization of human brain. Gray matter and gyrification describe the geometric characteristic of cortical surface and the wiring of white matter represents the structural pathway inside the brain. Many studies suggest that there exists a close relationship between gray matter and white matter. However, given the widely existing variability and complexity of brain structures, it is still largely unknown to what extent white matter wiring can influence gray matter and folding patterns. As an attempt to discover the potential relationship between gray matter and white matter, in this work we developed a novel spatial hybrid U-Net framework for multi-modal image prediction: we are aiming to predict T1-weighted Magnetic Resonance Imaging (MRI) based on Diffusion Tensor Imaging (DTI) data. Specifically, when predicting local intensity for T1 data, we constructed a hybrid model to integrate both local tensor information and the FA (Fractional Anisotropy) measure from remote brain regions connected by DTI derived fibers. To alleviate computation effort and reduce memory consumption, we proposed a multi-stage 2D training scheme instead of using 3D convolution neural network. Our results showed 80% accuracy for prediction and the reconstructed cortical surface using predicted T1 data is highly consistent to the original T1 derived surface. We envision that the proposed method can not only lay down a foundation for multi-modality inference, but also bring new insights to understand brain structure as well.

Keywords: Brain structure · Multi-modality · U-Net

1 Introduction

Human brains display significant inter-individual variation in cortical structures. Folding pattern of the cerebral cortex (gray matter - GM) and white matter (WM) connectivity are two aspects of brain structure. These two characteristics together compose the structural organization of human brain. That is, gray matter and gyrification depict the geometric shape of cortical surface, e.g., gyri and sulci, and white matter wiring provides the white matter pathway inside the cortex. Many studies suggest that there exists a close relationship between gray matter and white matter, such as a universal scaling law between GM and WM [1]. There are two divergent ideas in the field: axonal pushing and pulling [2] theory. A recent study further proved that during cortical gyrification, gyral regions with higher concentrations of growing axonal fibers tend to form 3-hinge

© Springer Nature Switzerland AG 2020
Q. Li et al. (Eds.): MMMI 2019, LNCS 11977, pp. 1–9, 2020.
https://doi.org/10.1007/978-3-030-37969-8_1

gyri [3]. All the above studies suggest that there exists a close relationship between gray matter and white matter. However, given the existing variability and complexity of brain structures, it is still largely unknown to what extent white matter wiring can influence gray matter and folding patterns.

Fortunately, the advancement of MRI and DTI provides non-invasive ways to examine cortical folding patterns and white matter related measures (e.g. FA maps and DTI derived fibers). As an attempt to discover the potential relationship between gray matter and white matter, we developed a novel spatial hybrid U-Net framework for multi-modal image prediction: we aim to predict T1-weighted MRI based on DTI data. The motivation is that if white matter influences cortical folding (gray matter) via either pulling/pushing or their combined effects, we should be able to infer gray matter with the information of white matter. Specifically, when predicting local intensity for T1-weighted image data, we constructed a hybrid model to integrate both local tensor information and the FA measures from remote brain regions connected by DTI derived fibers. The reason for integrating local and remote white matter knowledge as a hybrid model is to examine if structural connectivity will also contribute to local gray matter properties. To alleviate computation effort and reduce memory consumption, we proposed a multi-stage 2D training scheme instead of using 3D convolution neural network (CNN). Using Human Connectome Project [4] data set as a test bed, our results showed 80% accuracy for T1-weighted image prediction with DTI data. Here, the accuracy is defined as the ratio of predicted intensity to the original one. We also examined the reconstructed cortical surfaces with the predicted T1 data and it displayed high similarity to the surfaces generated from the original T1 image.

2 Method

2.1 Data Acquisition and Pre-processing

The data used in this work is acquired from the WU-Minn Human Connectome Project (HCP) consortium S1200 Release. We use T1-weighted MRI and DTI. For T1, TR = 2.4 s, TE = 2.14 ms, voxel size is 0.7 mm isotropic. For DTI, TR = 5.520 s, TE = 89.5 ms, slice thickness is 1.25 mm. The diffusion weighted data consists of 3 shells of b = 1000, 2000, and 3000 s/mm^2 with an approximately equal number of acquisitions on each shell within each run. In this work, we only used 90 b = 1000 volumes and 1 b = 0 volume.

The data acquired is then pre-processed through a series of steps using the FMRIB Software Library (FSL) [5, 6]. Finally, the T1-weighted images are registered to their respective DTI b = 0 images using FMRIB's Linear Image Registration Tool. [7, 8] This step allows us to register the DTI and FA images in the same space as the T1 image.

2.2 Extracting Features from Data

To alleviate computation effort and reduce memory consumption, we proposed a multi-stage 2D training scheme instead of using 3D CNN. That is, each data sample for our model training and prediction is generated based on a single 2D slice. Figure 1

demonstrates the construction of our input. The features contained in each sample comprise two parts: the first part comes from local structural information which includes the gradients of ninety b = 1000 volumes and one b = 0 volume. The second part of the input represents the remote structural information. For example, for each voxel considered, we can compute its connected voxels (the colored voxels in Fig. 1) in remote regions by examining if there are fibers passing through both of them. For the connected voxels, we acquire their FA measures and concatenate them with the local information as the entire input. Because of computation constrains, we considered 6 remote regions for each slice. Different features are treated as different channels. Totally we have 97 channels for each 2D slice. The motivation to integrate both local and remote structural information together for model training and prediction is to examine if the remote brain regions can influence the local brain structures via WM structural connectivity.

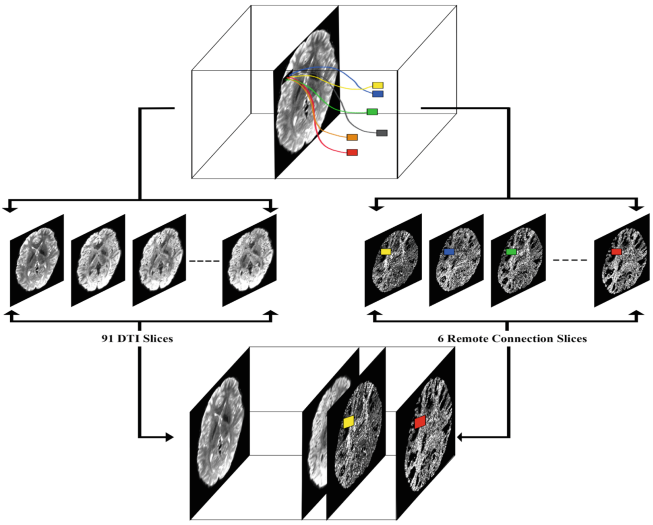

Fig. 1. Combining DTI and FA slices as features

2.3 Transfer Learning Using a Spatial Hybrid U-Net Model

The deep neural network architecture adopted in this work is similar to the U-Net model [9]. This model can be divided into two major parts with 5 blocks on the left (down-sampling) and 5 blocks on the right (up-sampling) of a central block. The first block of the model represents the input layer. This is a collection of 2D image slices from the 3D volume we generated from the DTI and FA images. This is followed by four blocks of similar structure (down sampling) - two 3×3 2D convolutions with a rectified linear unit (ReLU), followed by a 2×2 max pooling operation. Batch normalization and a dropout of 25% is used after each max pooling layer. Before we begin the up-sampling blocks, we apply two 3×3 convolutions with ReLU followed by batch normalization and a dropout of 50% nodes. The four blocks of the up-sampling

part consist of a concatenate operation to merge the layers in the same spatial order. This is the key feature of the U-Net architecture.

This is followed by two similar 3 × 3 2D convolution operations and batch normalization. The final up sampling block is passed through a 1 × 1 2D convolution with a single filter to regress the output. Note that even though the input of the proposed model is based on 2D slices, it contains the 3rd dimension of 97 channels which integrate both local and spatially remote structural information. Therefore, we named it as spatial hybrid U-Net. Figure 2 visualizes the overall structure we used in this adaptation of the U-Net.

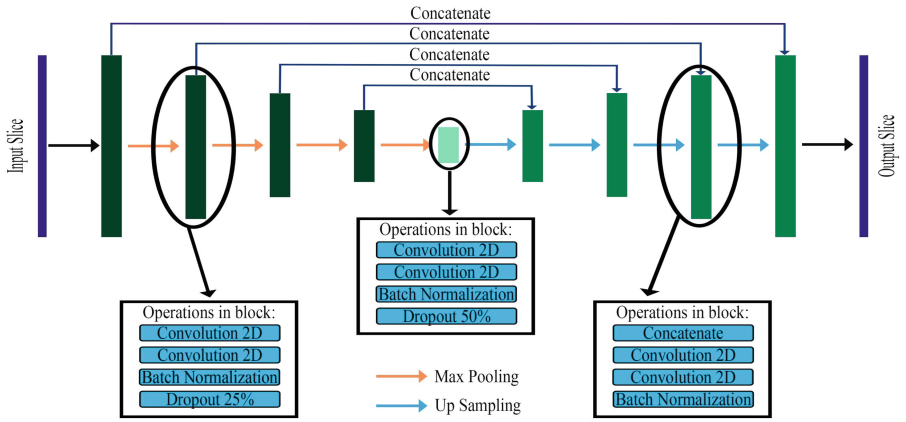

Fig. 2. Spatial hybrid U-Net architecture. Each color of block represents a different type of combination of operations. The input and output boxes are 2D slices images. One block from each type has been enlarged to show the operations that it contains.

2.4 Training and Prediction Procedure

In this work, we proposed a multi-stage training strategy: the entire training process is divided into three stages corresponding to three image directions. First, we train the model from the sagittal plane. For this, we use the image slices pre-processed with the sagittal direction and feed into our spatial hybrid U-net model. These image slices are from 15 subjects (subject 1 to 15) and each slice is treated as an independent training sample. This part of the training helps us achieve a baseline for the weight initialization, which will be used to train the model with images from other planes. The images from the axial plane are taken from a different set of 15 subjects (subject 16 to 30) and the model is continued to be trained from the axial direction. Finally, we train the model from the coronal direction using coronal slices from another set of 15 subjects (subject 31 to 45). The reason that we did not conduct the training using three directions simultaneously is that there exists significant difference of the training performance for different directions. Essentially, we trained the model from a direction with

the best prediction performance and used that as pre-knowledge to train the next one. This is similar to transfer learning that transfers the knowledge gained from one plane to a different plane. Figure 3 shows the overview of our training process.

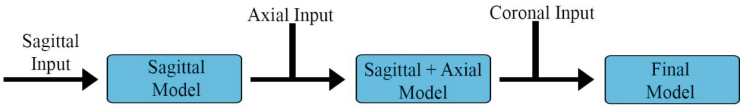

Fig. 3. Overview of training process

The optimizer used to train the model is the stochastic gradient-based Adam optimizer. It updates the model parameters with the rule:

$$\Omega^{t+1} = \Omega^{t} - \gamma . \frac{\widehat{m_a^t}}{\sqrt{\widehat{m_b^t}} + v} \tag{1}$$

The Adam optimizer computes network weights and biases Ω^{t+1} for step t + 1 using adaptive momentum estimation technique, which enables it to compute parameters using variable learning rate (γ) instead of a fixed constant learning rate. This enables the optimizer to calculate gradient steps more precisely enabling it to converge faster. It uses first and second exponential decay rates β_a *and* β_b to compute first and second momentum estimates $\widehat{m_a^t}$ and $\widehat{m_b^t}$. These values are used to compute new weights and biases for the network at t + 1 step. The parameter ν is a regularization parameter used to avoid division by zero cases.

We use Mean Squared Error to penalize the difference in the predicted value. The loss is calculated by the following formula:

$$\mathrm{f}(\Omega^{t}) = \frac{1}{N}\sum_{i=1}^{N}(t_i - p_i)^2 \tag{2}$$

Here, t_i *and* p_i are true and predicted values of the deep neural network and N is a normalizing constant defined by the number of subjects (number of training samples).

3 Results

3.1 Predicting T1-Weighted Images from DTI and FA Images

Given the 90 gradients extracted from 90 b = 1000 s/mm^2 DWI volumes (local information) as well as FA measures of voxels connected (remote information) for each voxel (Sect. 2.2), we conducted T1 prediction model training with spatial hybrid U-Net

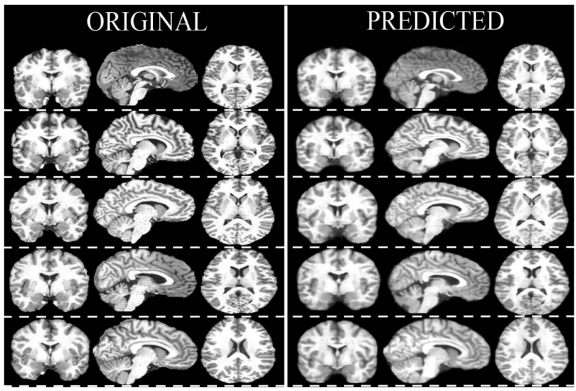

Fig. 4. Comparison of predicted T1-weighted image and original T1 data. We used 45 subjects for prediction model training (15 subjects for each stage). Here, we randomly select 5 subjects and show their prediction results (right) with the original T1 data (left) as well

frame-work proposed in Sects. 2.3 and 2.4. In this work, we perform the prediction (50 subjects) from all three directions and the final result is averaged. We randomly select 5 subjects and show their original T1 data as well as prediction results in Fig. 4. We can see that both overall structures and detailed spatial patterns of GM/WM are highly consistent between our prediction results and the original T1 data. The related quantitative results are shown in Sect. 3.3.

3.2 Predicting Cortical Surfaces with Predicted T1-Weighted Images

For the 50 predicted T1 images using our DTI prediction model, we conducted the cortical surface reconstruction as regular T1 images. Each sub-figure in Fig. 5 shows the reconstructed cortical surfaces using the original T1 image (top) and our predicted one (bottom). By visual examination we can see that the overall folding patterns are highly consistent between the original T1 derived cortical surfaces and the ones based on DTI prediction results. The color encodes the prediction error and the regions with red represent high prediction error. An interesting observation is that the temporal lobe (red circles) and dorsal regions (red arrows) tend to have higher prediction error compared to other brain regions. This might be due to the more complex structure of the corresponding regions. Importantly, we can accurately predict T1 image and generate cortical surface only using DTI data, which was considered a different imaging modality.

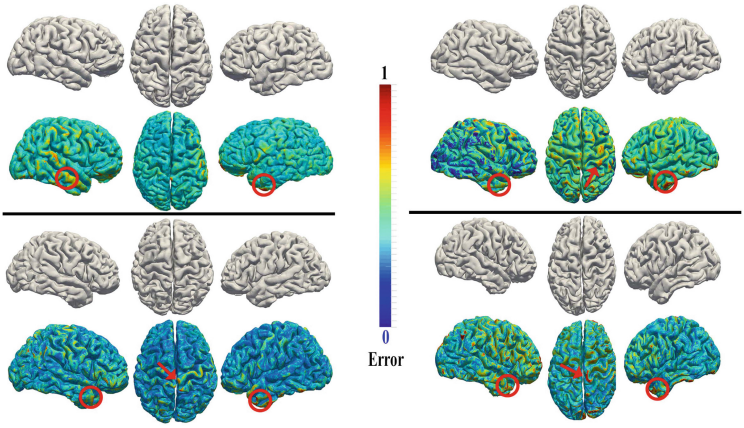

Fig. 5. Examples of cortical surface reconstruction using original and predicted T1 images (Color figure online)

3.3 Comparison with Model Without FA Slices in the Training Data

In our paper, we used Mean Absolute Error (MAE) to measure the quality of predicted T1-weighted images. MAE is a measure of difference between two continuous variables. The Mean Absolute Error is given by:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |a_i - b_i|}{\text{n}} \tag{3}$$

where a_i and b_i are the original T1 data and predicted T1 data, respectively.

Our comparison results can be seen in Table 1. 5 predicted subjects were randomly chosen and compared with their original registered T1-weighted image. It is evident that including remote features (FA) has a positive effect on the prediction model.

Table 1. Comparison of our proposed DTI + FA method with only DTI method (MAE values)

Subject	1	2	3	4	5	Average
DTI + FA method	0.073	0.072	0.074	0.057	0.073	**0.0714**
Only DTI method	0.081	0.079	0.082	0.069	0.084	0.0790

3.4 Comparison with k-Nearest Neighbor (kNN) Regression

We compared the results of our model with kNN Regression (n = 5) and it is evident from the results that our proposed model performs better. The MAE values have been compared to show the performance improvement that our model achieves (Table 2).

Table 2. Comparison of our proposed model with kNN regression (MAE values)

Subject	1	2	3	4	5	Average
kNN	0.120	0.112	0.115	0.109	0.130	0.1172
Proposed	0.073	0.054	0.057	0.063	0.068	**0.0630**

3.5 Reproducibility

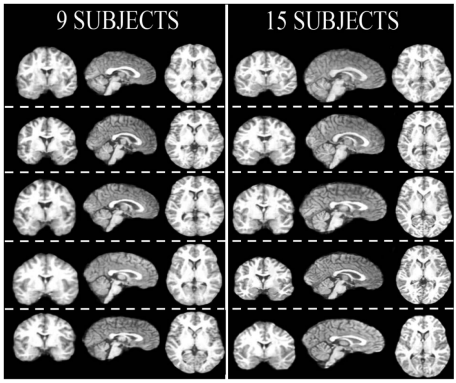

Fig. 6. Prediction using 9 subjects (left) and 15 subjects (right)

We examined the robustness of our model by using the same training procedure for 9 subjects (3 from each plane) and 15 subjects (5 from each plane). The results were satisfactory, as can be seen from Fig. 6 above. Using higher number of subjects yields higher accuracy, as expected. However, our model proves to be robust in terms of the T1-weighted predictions even with very few subjects used for training.

4 Conclusion

In this work, we developed a novel spatial hybrid U-Net model to predict T1 image from DTI data. During this process, we used both the local structural information and also FA measures from remote regions connected by DTI derived fibers. We also proposed a multi-stage training scheme to achieve a more reliable prediction performance. Our results showed 80% accuracy for prediction and the reconstructed cortical surface using predicted T1 data is highly consistent to the original T1 derived surface. We envision that the proposed method can not only lay down a foundation for multi-modality inference, but also bring new insights to understand brain structure as well.

References

1. Zhang, K., Sejnowski, T.J.: A universal scaling law between gray matter and white matter of cerebral cortex. Proc. Natl. Acad. Sci. U.S.A. **97**(10), 5621–5626 (2000)
2. Razavi, M.J., et al.: Radial structure scaffolds convolution patterns of developing cerebral cortex. Front. Comput. Neurosci. **11**, 76 (2017)
3. Ge, F., et al.: Denser growing fiber connections induce 3-hinge gyral folding. Cereb. Cortex **28**(3), 1064–1075 (2017)
4. Van Essen, D.C., et al.: The WU-Minn human connectome project: an overview. NeuroImage **80**, 62–79 (2013)
5. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. NeuroImage **62**, 782–790 (2012)
6. Smith, S.M., et al.: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage **23**(S1), 208–219 (2004)
7. Jenkinson, M., Smith, S.M.: A global optimisation method for robust affine registration of brain images. Med. Image Anal. **5**(2), 143–156 (2001)
8. Jenkinson, M., Bannister, P.R., Brady, J.M., Smith, S.M.: Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage **17**(2), 825–841 (2002)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28