Transfer Learning from Partial Annotations for Whole Brain Segmentation

Chengliang Dai[1(✉)], Yuanhan Mo[1], Elsa Angelini[2], Yike Guo[1],
and Wenjia Bai[1,3]

[1] Data Science Institute, Imperial College London, London, UK
c.dai@imperial.ac.uk
[2] ITMAT Data Science Group, Imperial College London, London, UK
[3] Department of Brain Sciences, Imperial College London, London, UK

Abstract. Brain MR image segmentation is a key task in neuroimaging studies. It is commonly conducted using standard computational tools, such as FSL, SPM, multi-atlas segmentation etc, which are often registration-based and suffer from expensive computation cost. Recently, there is an increased interest using deep neural networks for brain image segmentation, which have demonstrated advantages in both speed and performance. However, neural networks-based approaches normally require a large amount of manual annotations for optimising the massive amount of network parameters. For 3D networks used in volumetric image segmentation, this has become a particular challenge, as a 3D network consists of many more parameters compared to its 2D counterpart. Manual annotation of 3D brain images is extremely time-consuming and requires extensive involvement of trained experts. To address the challenge with limited manual annotations, here we propose a novel multi-task learning framework for brain image segmentation, which utilises a large amount of automatically generated partial annotations together with a small set of manually created full annotations for network training. Our method yields a high performance comparable to state-of-the-art methods for whole brain segmentation.

1 Introduction

Magnetic resonance imaging (MRI) plays an important role in human brain studies due to its good performance on presenting anatomy, pathology and function of the brain. Accurate segmentation of brain MRI scans is a prerequisite for measuring volume, thickness and shape of brain structure, which allows researchers to track and study the development, ageing and diseases of the brain [1]. Brain image segmentation is a time-consuming process when conducted manually, which typically takes several hours for a single subject. Therefore computational tools including FSL [2], SPM [3], MALP-EM [4] etc have been developed to automatically segment brain MRI scans and to enable large-scale population-based

imaging studies. Most of these computational tools segment the scans by performing linear and nonlinear registration between a manually annotated brain atlas and a target scan and then propagating the atlas. Despite the efficiency they bring, these tools still suffer problems such as expensive computational cost and potential failures in image registration. Furthermore, strict pre-processing steps including brain stripping and bias correction are required to improve the reliability of these computational tools.

Neural networks have been explored and widely used for brain segmentation in recent years. Comparing to conventional brain image segmentation pipelines that are registration-based, network-based methods use pairs of images and manual annotations to train a discriminative model for inferring the segmentation of a new scan. Such difference brings a few advantages: (i) pre-processing can be potentially simplified [5]; (ii) processing time is significantly reduced without sacrificing the segmentation accuracy. Segmenting brain with network-based models also has drawbacks as these models require massive amount of annotated data for model training. The limited amount of annotations for brain images has become one of the biggest challenges for applying neural networks to brain image segmentation.

Previous works have been exploring ways for training image segmentation networks with limited annotations. A common approach is to fine-tune a pre-trained network from large image datasets like ImageNet [6]. In [7], an encoder-decoder model is pre-trained with auxiliary labels generated by FreeSurfer and then fine-tuned with an error corrective boosting loss. In [8], a multi-task image segmentation model is investigated to learn features that can be shared between MRI scans of different parts of human body. Generative adversarial networks (GANs) are adopted in [9] for data augmentation, which indicates a better performance than conventional augmentation methods.

Here we propose a novel brain image segmentation network, which leverages a massive set of automatically generated partial annotations (sub-cortical segmentations from FSL) for network pre-training and then perform transfer learning onto a small set of full annotations (manual whole brain segmentations). Compared to [7], our method is conducted in 3D but with less convolutional
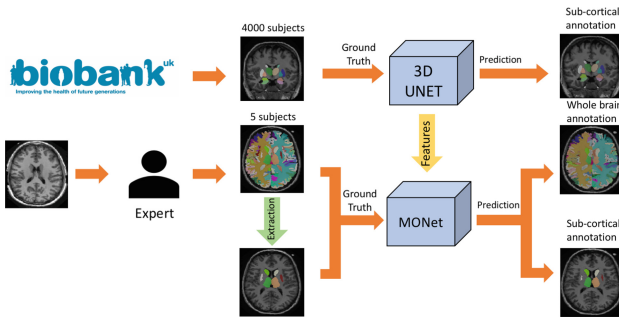

Fig. 1. Two-stage training scheme: Stage 1: pre-training; Stage 2: joint training.

layers. We demonstrate how features learnt from partial annotations in the source domain can be adapted to the target domain. With very limited annotations, our method achieves a performance comparable to state-of-the-art methods for brain image segmentation.

2   Method

Our work adopts a two-stage training scheme as illustrated in Fig. 1. Stage 1 pre-trains the segmentation network using a large set of automatically generated partial annotations. Stage 2 fine-tunes the network by jointly training on partial annotations and a small set of full annotations.

2.1   Pre-training with Partial Annotations

In this work, partial annotation refers to segmentation that only covers part of the brain structures. In our case, it refers to segmentation of 15 sub-cortical structures automatically generated by FSL. Full annotation refers to segmentation of whole brain structures manually annotated by human experts, which is a superset of partial annotation and consists of 138 structures. Since partial annotations are automatically generated, it is easy to acquire many of them. On the other hand, acquiring full annotations is more difficult as it requires extensive manual labour.

A 3D U-Net is employed for pre-training on partial annotations, using categorical cross-entropy as the loss function,

$$\mathcal{L} = -\sum_{v} g_l^w(v) \log p_l^w(v) \tag{1}$$

where $p_l^w(v)$ is the the predictive probability of partial segmentation belonging to class l at voxel v and $g^w(v)$ is the probability of it belonging to its actual class.
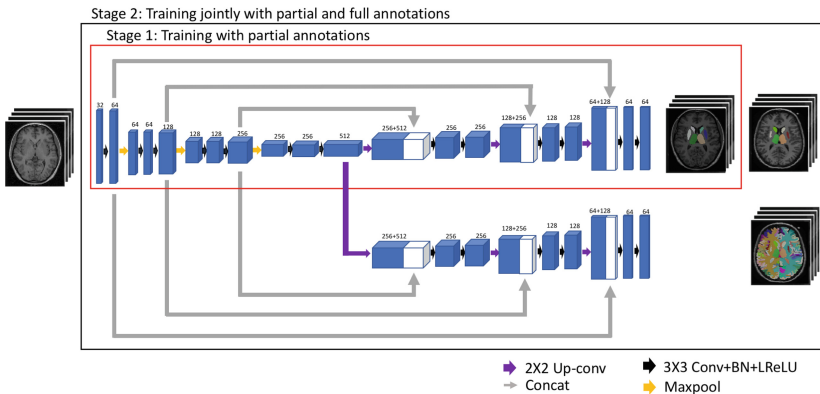

Fig. 2. Network architectures used in stage one and two training.

2.2   Joint Training with Full Annotations

We employ a multi-task learning framework for the second stage. The encoder is consistent with the architecture used in the first stage. Two decoders are used, so that the two tasks (partial segmentation and full segmentation) can be jointly trained. We refer our method as the multi-output network (MO-Net). The encoder and both decoders are loaded with the pre-trained parameters. Multi-output design encourages the encoder to learn shared features for partial segmentation and full segmentation. The partial segmentation used for joint training is extracted from the full segmentation, which are manual segmentations of the whole brain. Since manual segmentations have always been considered as 'gold standard' and should be more reliable than segmentations from automatic tools, the trained MO-Net should also be able to provide more accurate partial segmentation than the one trained in the first stage. The multi-output design given in Fig. 2 is similar to the one described in [5], which allows the network to learn jointly from two segmentation maps in order to achieve more accurate prediction and to have the potential to provide segmentation output for various annotation protocols. However, the difference is that we use a modified U-Net instead of ResNet and FCN adopted in [5], and our network is loaded with the parameters learnt from the pre-training stage.

A weighted loss that combines the overall loss of two decoders of MO-Net for joint training is formulated as,

$$\mathcal{L}_{MO-Net} = -\sum_v \lambda_s g^s(v) \log p_m^s(v) - \sum_v \lambda_w g^w(v) \log p_l^w(v) \qquad (2)$$

where $p_m^s(v)$ is the predictive probability of full segmentation belonging to class m at voxel v and $g^s(v)$ is the probability of it belonging to its actual class. λ_s and λ_w are the weights for overall loss function. To balance between the learning tasks for partial segmentation and full segmentation, we assign 0.5 to both losses in the overall loss function.

3   Experiments and Results

3.1   Datasets

UK Biobank Dataset (UKBB). 4,000 MRI brain scans from the UK Biobank are used. Automatic sub-cortical segmentations of 15 regions by FSL are used as partial annotations for pre-training.

Hammers Adult Atlases (HAA). The HAA dataset [10,11] contains brain atlases for 20 subjects with manual annotations for 67 regions. The dataset is split into 5/2/13 for training, validation and test.

MICCAI 2012 Multi-atlas Labelling Challenge (MALC). The MALC dataset [12] contains MRI scans from 30 subjects (15 subjects for training) with manual annotations for the whole brain for 138 regions and 132 regions are used

for performance evaluation. The dataset also includes 5 follow-up scans, but they are excluded in our work. The dataset is split into 15/2/13 for training, validation and testing.

The manual annotations from the HAA and MALC datasets are regarded as ground truth in evaluation.

3.2   Preprocessing and Training

The typical brain image resolution is 256^3, with isotropic spatial resolution of $1\,\mathrm{mm}^3$. All images were rigidly registered to MNI space and normalized to zero mean and unit standard deviation. For training the network, 3D patches of size 128^3 were randomly drawn from the brain images. Batch size was set to 1 due to the limitation of GPU memory. Random elastic deformation was applied to the 3D patches for data augmentation. Cropping and augmentation were performed on-the-fly. Adam optimiser with a starting learning rate of 0.001 was used for both stages of network training. Leaky rectified linear unit (LeakyReLU) with a negative slope of 0.01 is applied as the activation function. For the proposed method, pre-training was ran for 3 epochs and joint training was ran for 200 epochs. We also trained a standard U-Net as a baseline method for comparison.

3.3   Results

We evaluated the performance of MO-Net in terms of Dice score. For comparison, two versions of U-Nets were trained, one trained from scratch (U-Net (FS)) and the other fine-tuned (U-Net (FT)) on MALC and HAA respectively. For evaluating whole brain segmentation performance on MALC, we also compared our result to SLANT8 and SLANT27 [13], which is based on fine-tuning 8 and 27 3D U-Nets pre-trained with 5111 subjects for different locations of brain.

As shown in Tables 1 and 2, our method outperformed the U-Net trained from scratch by 26% on MALC dataset and 19% on HAA dataset. MO-Net also shows slight improvements over fine-tuned U-Net, SLANT8 and SLANT27 on both MALC and HAA datasets. We further compared to QuickNAT [7] on the same 25 brain structures as in their paper on the MALC dataset. The result is given in Table 3. MO-Net outperformed the fine-tuned U-Net, SLANT8 and SLANT27 by a small margin, although the performance is inferior to QuickNAT.

Table 1. Whole brain segmentation accuracy on MALC.

Method	Dice (mean $\pm$ std)
U-Net (FS)	0.623 ± 0.095
U-Net (FT)	0.782 ± 0.043
SLANT8 [13]	0.768 ± 0.011
SLANT27 [13]	0.776 ± 0.012
MO-Net	$\mathbf{0.785 \pm 0.070}$

Table 2. Whole brain segmentation accuracy on HAA.

Method	Dice (mean $\pm$ std)
U-Net (FS)	0.706 ± 0.032
U-Net (FT)	0.821 ± 0.019
MO-Net	$\mathbf{0.843 \pm 0.037}$

Table 3. Segmentation accuracy for 25 structures on MALC.

Method	Dice (mean $\pm$ std)
U-Net (FS)	0.775 ± 0.035
U-Net (FT)	0.809 ± 0.021
SLANT9 [13]	0.817 ± 0.036
SLANT27 [13]	0.823 ± 0.037
QuickNAT [7]	0.901 ± 0.045
MO-Net	0.838 ± 0.049

Table 4. Segmentation accuracy for 15 sub-cortical structures on MALC.

Method	Dice (mean $\pm$ std)
U-Net (FS)	0.649 ± 0.145
U-Net (FT)	0.835 ± 0.062
FSL	$0.637(9 \text{ failed}) \pm 0.216$
MO-Net	0.826 ± 0.029

Table 5. Segmentation accuracy for 15 sub-cortical structures on HAA.

Method	Dice (mean $\pm$ std)
U-Net (FS)	0.612 ± 0.103
U-Net (FT)	0.874 ± 0.053
FSL	0.763 ± 0.043
MO-Net	$\mathbf{0.879 \pm 0.091}$

For sub-cortical segmentation, we compared our result to FSL FIRST and U-Net. The proposed method MO-Net shows similar Dice score performance to fine-tuned U-Net and it is better than FSL and U-Net trained from scratch. The result is shown in Tables 1 and 4.

A box-plot of Dice scores comparing MO-Net with U-Net trained from scratch and fine-tuned on HAA for 8 brain structures is given in Fig. 3 showing the improvement of adopting our method. A qualitative result of whole brain and sub-cortical segmentation from MO-Net is given in Fig. 4, which shows better segmentation accuracy for certain structures comparing with U-Net and FSL (Table 5).

The result has demonstrated that a CNN-based model pre-trained with partial segmentation can achieve better accuracy for whole brain segmentation. The performance of MO-Net in terms of Dice scores is comparable to 3D U-Net based approaches in [13] on MALC with less strict training data, although inferior to [7] probably due to the deeper network they adopted. We believe the performance of our approach has the potential to be improved with a more advanced CNN design in the future. In general, multi-task learning helps the model to improve the generalization and in our case, to learn features shared by partial segmentation and full segmentation, which can possibly make our encoder more robust. Such claim would need more experiments to prove in the future.
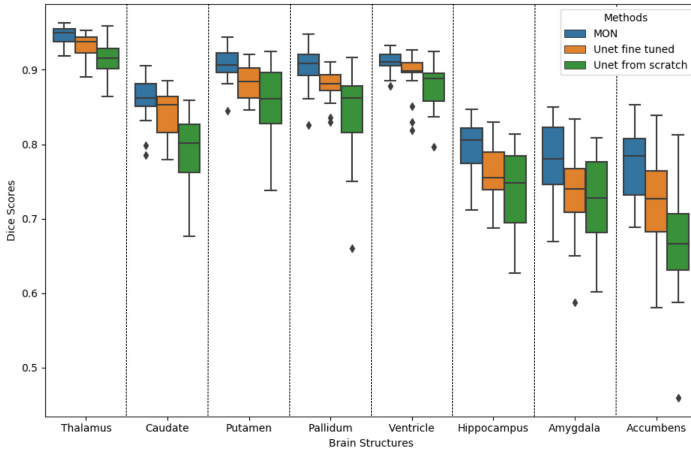
Fig. 3. Box-plot of Dice scores of MO-Net, U-Net fine-tuned and U-Net trained from scratch on HAA for 8 brain structures on the left hemisphere.
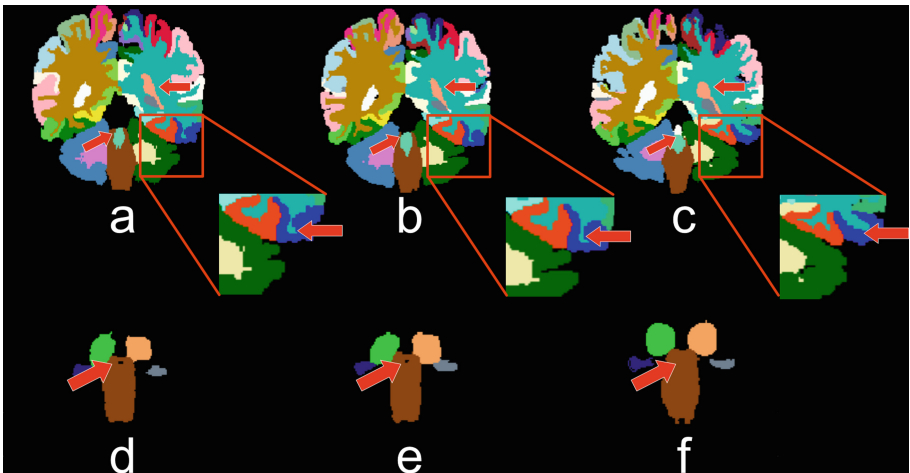

Fig. 4. Visual inspection of whole brain segmentation and sub-cortical segmentation on MALC: Ground truth of full (a) and partial (d) brain segmentation from the expert, full (b) and partial (e) brain segmentation from MO-Net, full (c) segmentation from fine-tuned U-Net, and sub-cortical (f) segmentation from FSL. Red arrows indict regions where MO-Net looks consistent with manual annotations and outperforms other methods. (Color figure online)

4   Conclusion

In this paper, we propose a method that combines transfer learning and multi-task learning to address the small data learning problem. Our method takes

advantage of existing automatic tool to create a large set of partial annotations for model pre-training which has been demonstrated to improve segmentation accuracy. The preliminary result on whole brain segmentation shows a good potential of the proposed method.

References

1. Douaud, G., Groves, A.R., Tamnes, C.K., Westlye, L.T., Duff, E.P., et al.: A common brain network links development, aging, and vulnerability to disease. Proc. Natl. Acad. Sci. **111**(49), 17648–17653 (2014)
2. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. NeuroImage **62**(2), 782–790 (2012)
3. Ashburner, J., Friston, K.J.: Voxel-based morphometry - the methods. NeuroImage **11**(6), 805–821 (2000)
4. Ledig, C., et al.: Robust whole-brain segmentation: application to traumatic brain injury. Med. Image Anal. **21**(1), 40–58 (2015)
5. Rajchl, M., Pawlowski, N., Rueckert, D., Matthews, P.M., Glocker, B.: NeuroNet: fast and robust reproduction of multiple brain image segmentation pipelines. In: International Conference on Medical Imaging with Deep Learning (MIDL) (2018)
6. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
7. Roy, A.G., Conjeti, S., Navab, N., Wachinger, C., Alzheimer's Disease Neuroimaging Initiative: QuickNAT: a fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage **186**, 713–727 (2019)
8. Moeskops, P., et al.: Deep learning for multi-task medical image segmentation in multiple modalities. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 478–486 (2016)
9. Shin, H.C., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International Workshop on Simulation and Synthesis in Medical Imaging, pp. 1–11 (2018)
10. Hammers, A., et al.: Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. **19**(4), 224–247 (2003)
11. Gousias, I.S., et al.: Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. NeuroImage **40**(2), 672–684 (2008)
12. Landman, B., Warfield, S.: MICCAI 2012 workshop on multi-atlas labeling. In: Medical Image Computing and Computer Assisted Intervention Conference (2012)
13. Huo, Y., et al.: 3D whole brain segmentation using spatially localized atlas network tiles. NeuroImage **194**, 105–119 (2019)