Training Data Requirements for Atlas-Based Brain Fibre Tract Identification

Fiona Young[1*], Kristian Aquilina[2], Chris A. Clark[1], Jonathan D. Clayden[1]

***Abstract*— Large volumes of annotated training data are often required for data-driven image analysis methods. We consider two techniques for identifying brain fibre bundles from diffusion MRI scans, tractfinder and TractSeg, and compare performances using different amounts of training data. Our results show that tractfinder, an atlas-based method, shows no improvement in performance beyond a relatively small number of training samples. This is an advantage in a field where generating and maintaining high quality reference data is difficult and time-consuming.**

***Keywords*—medical imaging, deep learning, diffusion MRI, segmentation**

I. Introduction

In segmentation methods relying on seen data from which to "learn" patterns to recognise in unseen data, the volume and range of training data influences the prediction accuracy and generalisability. Deep learning models, which have many thousands of network parameters to learn, can require immense amounts of training data, posing a particular barrier to the use of such models in applications where suitably annotated data is scarce.

One such area is the problem of segmenting white matter fibre tracts, structures which form the communication pathways between different centres of the brain, from diffusion weighted magnetic resonance imaging (dMRI) data [1]. There remains no means for obtaining ground truth information for this task: the complex arrangement of fibres can only be indirectly probed through dMRI on a millimetre scale. The global anatomies of tracts are only inferred from a combination of post-mortem dissections, functional observations and lesion studies, methods which are themselves far from definitive [2]. New pathways are regularly proposed and existing known connections continually revisited [2]. The effort involved in creating accurate reference annotations for the task of segmenting fibre tracts in dMR images is substantial [1], and one at risk of requiring duplication due to the evolving nature of the problem. Atlas-based methods could provide less data-intensive solutions with comparable results, while one-shot or few-shot learning has also been explored for generalising pre-trained deep learning models to new tracts using only few additional training samples [3]. We set out to consider the volume of training data required by two different proposed tract segmentation techniques: a deep learning and an atlas-based approach.

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4health) (EP/S021930/1) and the Department of Health's NIHR-funded Biomedical Research Centre at Great Ormond Street Hospitals.

[1]Department of Developmental Neurosciences, Institute of Child Health, University College London, London, United Kingdom

[2]Department of Neurosurgery, Great Ormond Street Hospital, London, United Kingdom

fiona.young.15@ucl.ac.uk

II. Description of segmentation methods

A. Tractfinder

Tractfinder [4] is an atlas-based method for identifying tracts in dMRI brain scans. It relies on a tract orientation distribution (TOD) atlas to encode the expected spatial distribution and expected local orientations of the bundle. Diffusion MRI data of the target image is preprocessed and the local fibre orientation distributions (FODs) estimated using constrained spherical deconvolution (CSD) [5]. The atlas is then registered to and compared with the target data by taking the inner product of the two spherical distributions. The number of subjects used to construct each tract atlas affects the amount of inter-subject anatomical variation reflected in the spatial and orientational expectations. We expect that the additional information to be gained from adding more training subjects would reach a point of saturation.

B. TractSeg

TractSeg [6] is a deep learning tract segmentation model which produces volumetric segmentations for 72 tracts directly from FOD peak directions. The default model was trained on streamline tractography bundle reconstructions as described in [6]. 105 subjects in total were used for cross-validation training, with the final trained model having seen 63 unique subjects. The effective volume of training data was also increased with various data augmentation techniques (see [6] for details). In [3], the TractSeg architecture was expanded to one-shot learning of novel tracts.

III. Experiments

We investigated the effect of training data volume on tractfinder segmentation performance by varying the number of subjects used to construct the atlases. To enable a direct comparison between the two methods, we used the same datasets and reference bundles used in the cross-validation of TractSeg, which are publicly available [7]. Using the same train–test data split as described in [7], subsets of 1, 3, 5, 10, 15, 30 and 63 training subjects were randomly selected, from which separate tractfinder TOD atlases where constructed. Tract segmentations were then obtained in the 42 test subjects using each of the different subset atlases and compared with the reference segmentations using the Dice similarity score (DSC). TractSeg segmentations were also generated for the same 42 test subjects. We limited this

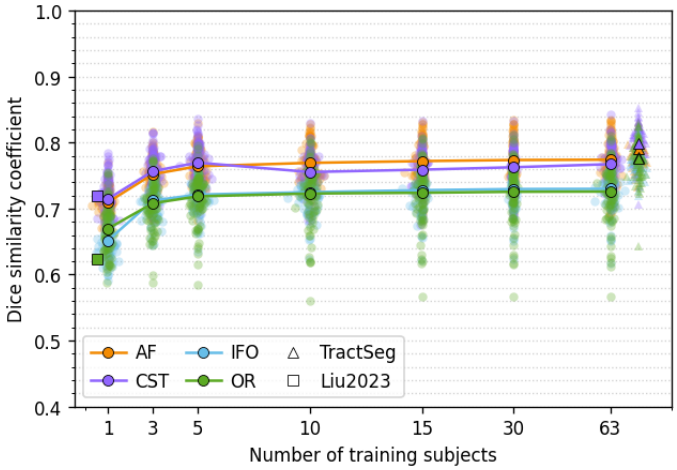

Fig. 1. Segmentation performance by number of atlas / model training subjects and tract. Individual subject data-points (mean across hemispheres) are represented by pale dots, mean across all subjects is indicated by dots with black borders. Relevant results from [3] are plotted as squares. AF = arcuate fasciculus; CST = corticospinal tract; IFO = inferior fronto-occipital fasciculus; OR = optic radiation.

analysis to the four tracts most commonly segmented for clinical purposes (e.g. for neurosurgical planning): arcuate fasciculus (AF), corticospinal tract (CST), inferior fronto-occipital fasciculus (IFO) and optic radiation (OR). The described methods involving de-identified human dMRI data were approved by the institutional research ethics committee.

IV. Results

When using only a single subject's normalised TOD map as an "atlas", mean DSC ranged from 0.65 to 0.71 for the IFOF and CST respectively. The maximum increase in mean DSC between the 15 and 63 subjects atlases was only 0.00835, for the CST, representing a 1% increase from the lower score of 0.759. Across all tracts, differences in performance became negligible beyond around 10 training subjects (Fig. 1). Overall, these results support the original use of 16 training subjects for tractfinder, as described in [4].

Scores for TractSeg were higher than tractfinder, but only appreciably so for the OR and IFO. The difference in mean DSC between TractSeg and the best tractfinder result was highest for the OR at 0.051 points (7% increase) and lowest for the AF at 0.016 points (2% increase). For the OR, tractfinder with one training subject performed better than the best one-shot learning model presented in [3] (Fig. 1).

V. Discussion

Increasing the number of atlas subjects beyond a minimum of around 10 to 15 does little to nothing to improve tractfinder results. Further training subjects offer minimal additional information on inter-subject variability, much of which is already smoothed out due to affine co-registration of training subjects into template space. While a perceived general attitude of "the more data the better" prevails in the machine learning space, we note that this doesn't necessarily hold for all data-driven image analysis techniques. This is of particular advantage in clinical applications, where novel tracts of interest to neurosurgeons are emerging, and where there is little capacity for generating large volumes of training data [3].

Training a new TractSeg model with each subset of the training data was not feasible for this study, so TractSeg results are only available for the full set of 63 training subjects. No motivation for the number of training samples is indicated in [6], and given the difficulty of creating reference annotations, we will assume that the creators of TractSeg considered this to be around the minimum required to achieve the desired accuracy. Nevertheless, the lack of direct comparison with TractSeg trained on fewer samples is a limitation of this study.

Deep learning and other data-heavy techniques are set to bring great advances to medical imaging. There is no denying the impressive accuracy achievable with deep learning segmentation as demonstrated by the TractSeg model, which our results show is not only more accurate but also more consistent, based on the reduced spread across all subjects. This strong performance, however, comes at the cost of producing a high number of high quality training data.

The need for large volumes of accurately annotated data may be easily fulfilled when the annotation process is straightforwards, or when resources are abundant. If not easy, the upfront effort is certainly justified if the application is a well-defined problem unlikely to need revisiting. Tract segmentation is none of those things: producing the ground truth reference annotated data is burdensome, and the likelihood that the effort may need duplicating as our understanding of white matter anatomy evolves is high. It is therefore worth asking whether marginal improvements in segmentation accuracy, as measured by the Dice score, are always worth the cost of producing the necessary volume of training data, und subsequent inflexibility of trained models.

References

[1] P. Poulin, G. Theaud, F. Rheault, E. St-Onge, A. Bore, E. Renauld, L. de Beaumont, S. Guay, P.-M. Jodoin, and M. Descoteaux, "TractoInferno - A large-scale, open-source, multi-site database for machine learning dMRI tractography," vol. 9, no. 1, p. 725. [Online]. Available: https://www.nature.com/articles/s41597-022-01833-1

[2] N. L. Voets, A. Bartsch, and P. Plaha, "Brain white matter fibre tracts: a review of functional neuro-oncological relevance," vol. 88, no. 12, pp. 1017–1025. [Online]. Available: http://dx.doi.org/10.1136/jnnp-2017-316170

[3] W. Liu, Z. Zhuo, Y. Liu, and C. Ye, "One-shot segmentation of novel white matter tracts via extensive data augmentation and adaptive knowledge transfer," vol. 90, p. 102968. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841523002281

[4] F. Young, K. Aquilina, C. A. Clark, and J. D. Clayden, "Fibre tract segmentation for intraoperative diffusion MRI in neurosurgical patients using tract-specific orientation atlas and tumour deformation modelling," vol. 17. [Online]. Available: https://doi.org/10.1007/s11548-022-02617-z

[5] J. D. Tournier, F. Calamante, and A. Connelly, "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution," vol. 35, no. 4, pp. 1459–1472.

[6] J. Wasserthal, P. Neher, and K. H. Maier-Hein, "TractSeg - Fast and accurate white matter tract segmentation," vol. 183, pp. 239–253.

[7] W. Jakob, N. Peter, and M.-H. Klaus, "High quality white matter reference tracts," Nov. 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1477956