Generalising the HCP multimodal cortical parcellation to UK Biobank

Logan Z. J. Williams[1,2], Matthew F. Glasser[3,4], Fidel Alfaro-Almagro[5], Simon Dahan[1], Abdulah Fawaz[1], Timothy S. Coalson[4], Sean P. Fitzgibbon[5], Mohamed A. Suliman[2], David C. Van Essen[4], Stephen M. Smith[5], A. David Edwards[1], and Emma C. Robinson[1,2]

[1] Department of Biomedical Engineering & Imaging Science, King's College London
[2] Centre for the Developing Brain, King's College London
[3] Department of Radiology, Washington University in St. Louis
[4] Department of Neuroscience, Washington University in St. Louis
[5] Centre for Functional Magnetic Resonance of the Brain, Oxford University

Abstract. The Human Connectome Project Multimodal Parcellation (HCP_MMP1.0) provides a robust *in vivo* map of the cerebral cortex, which demonstrates variability in structure and function that cannot be captured through diffeomorphic image registration alone. The HCP successfully employed a fully-connected neural network architecture to capture this variation, however it is unclear whether this approach generalises to other datasets with less rich imaging protocols. In this paper we propose and validate a novel geometric deep learning framework for generating individualised HCP_MMP1.0 parcellations in UK Biobank (UKB), an extremely rich resource that has led to numerous breakthroughs in neuroscience. To address substantial differences in image acquisition (for example, 6 minutes of resting-state fMRI per subject for UKB vs. 60 minutes per subject for HCP), we introduce a multi-step learning procedure including pretraining with a convolutional autoencoder. Compared to a fully-connected baseline, our proposed framework improved average detection rate across all areas by 10.4% (99.1% vs 88.7%), and detection of the worst performing area by 51.0% (80.9% vs. 29.9%). Importantly, this was not a result of the framework predicting one consistent parcellation across subjects, as visual inspection indicated that our method was sensitive to atypical cortical topographies. Code and trained models will be made available at `https://www.github.com`.

Keywords: Cortical parcellation · Geometric deep learning · UK Biobank

1 Introduction

The Human Connectome Project Multimodal Parcellation (HCP_MMP1.0) provides a robust *in vivo* map of the human cerebral cortex [9,10], which has been implicated in cognition, behaviour and neuropsychiatric disorders [25]. Using an observer-independent approach and multimodal cortical surface registration,
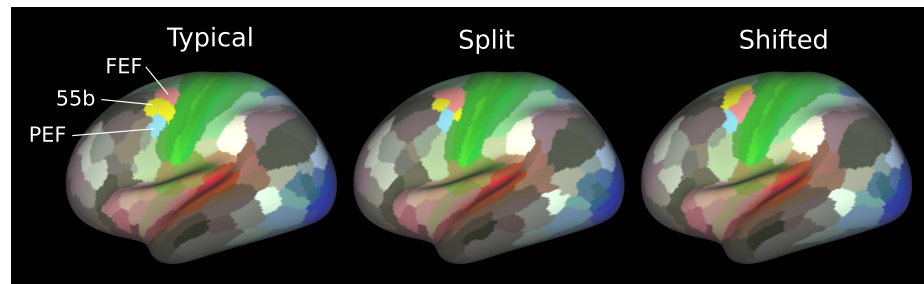
Fig. 1. Representative examples of typical, split and shifted area 55b topographies in individual HCP subjects, generated using the proposed geometric deep learning framework. Inter-individual differences reflect variabilities that persist following diffeomorphic multimodal alignment to a single group average template.

the HCP identified 83 areas previously reported using post-mortem histology, and 97 new areas defined by sharp transitions in structure and function [9]. Importantly, these cortical areas demonstrated marked diversity in structural and functional organisation that cannot be captured through biomechanically-constrained, diffeomorphic registration alone (Figure 1) [9]. To generate parcellations that reflected individual differences in cortical organisation, the HCP used 112 unique features of cortical structure and function to train a fully-connected neural network (FCNN) that accurately predicted which cortical area each vertex belonged to (360 networks in total), even when they significantly differed from the group average. These individualised HCP_MMP1.0 parcellations, which have significantly advanced our understanding of brain-behaviour relationships [4] and cognitive function [2], could benefit other large-scale neuroimaging consortia. One such study is the UK Biobank (UKB), an extremely rich resource that has led to a number of neuroscientific breakthroughs [18,7,26]. However, as these parcellations utilised large amounts of high-resolution MRI data per subject (including an extensive task fMRI battery) [10], it is unknown whether individualised, high-quality parcellations are also achievable in studies that cannot match these protocols [10,3].

The HCP FCNN approach demonstrated overall excellent performance, detecting 96.6% of all cortical areas in 210 unseen subjects. However, some areas were detected <75% of the time [9]. It is unknown whether these areas are truly absent, or whether this reflects inherent limitations in FCNN architectures, namely that each vertex is treated as spatially independent of its neighbours. Convolutional neural networks are one of the most successful learning-based methods for semantic segmentation, and operate by passing localised filters over an image to capture context-specific information [15]. This work introduces and validates a new geometric deep learning (gDL) framework for generating individualised HCP_MMP1.0 parcellations on sphericalised cortical surfaces. Specifically, this approach: 1) improves areal detection rates compared to a FCNN baseline; 2) remains sensitive to individual topographic variations in areal or-

ganisation that cannot be captured through diffeomorphic image registration alone; and 3) generalises the HCP_MMP1.0 parcellation to a new dataset with a substantially different image acquisition protocol.

2   Materials and Methods

Data HCP images were acquired with a customized 3T Siemens 'Connectom' Skyra scanner and standard Siemens 32-channel head coil at a single site in Washington University, St. Louis [11]. Training examples were taken from the same 210 HCP subjects ('210P'; 29.4 ± 3.5 years, 130 biological females (61.9%)) used to develop the original HCP MMP_1.0 [9]. HCP image acquisition consisted of structural: 0.7mm isotropic T1w and T2w; resting-state fMRI (rs-fMRI): 2.0mm isotropic, 60 minutes total; task-fMRI: 7 tasks, 48 minutes total [11].

Two cohorts were used for validation: 1) a separate set of 210 HCP subjects ('210V'; 28.8 ± 3.5 years, 114 biological females (54.3%), same as used by [9]) and 2) 1500 UKB subjects (63.8 ± 7.6 years, 822 biological females (54.8%)). UKB images were acquired with a standard 3T Siemens Skyra scanner and Siemens 32-channel head coil at 4 sites across England [1]. UKB image acquisition consisted of structural: 1mm isotropic T1w and T2-FLAIR; rs-fMRI: 2.4mm isotropic, 6 minutes total; task-fMRI: 1 task, 4 minutes total [1].

Pre-processing Major differences in scanning protocols exist between HCP and UKB datasets, with UKB acquiring lower resolution structural and functional imaging, and substantially less fMRI per subject (fewer tasks and 6 vs. 60 mins of rs-fMRI). This necessitated the development and validation of new tools for generalising the HCP_MMP1.0 to UKB data. To achieve this we first ran the HCP minimal preprocessing pipeline [11] to allow multimodal surface-based registration [22,21] of individual subject cortical features to the HCP MSMAll template space. This drives alignment using features on sphericalised cortical surfaces that better correspond with cortical areas, specifically: 1 T1w/T2w ratio ('myelin') map, 32 rs-fMRI ICA spatial maps and 9 visuotopic spatial maps) [9]. This required generating T1w/T2w ratio maps using the intensity bias correction described in [11], mapping the T1w/T2w ratio and denoised rs-fMRI timeseries to the cortical surface, and generating individualised rs-fMRI ICA spatial maps through weighted dual regression of the HCP group ICA spatial template[6] into UKB rs-fMRI timeseries data [9]. Compared to volumetric and folding-based surface registration, this multimodal surface registration approach markedly improves inter-individual alignment of cortical areas [9,5,22,21].

---

[6] Generating UKB specific ICA maps for multimodal alignment was considered, but was ultimately decided against in order to minimise differences in preprocessing with the HCP. However, group ICA of 3000 UKB subjects at dimension = 40 generated very similar components to group ICA of the same dimensionality in the HCP.

Model design Parcellation of the cerebral cortex was framed as a binary classification problem, where models were tasked with labelling vertices as being part of an area or not [9]. Model inputs were 91 cortical features including cortical thickness, curvature, T1w/T2w ratio, 77 rs-fMRI ICA spatial maps, 6 visuotopic spatial maps and 5 artefact maps[7]. Vertex classification was spatially constrained to occur within a restricted region (a 'searchlight' containing vertices within 30mm geodesic distance of the group average area). This was possible because all data had been multimodally aligned during pre-processing and justified since the relative positions of cortical areas are strongly conserved [14]. In total 360 classification networks were trained, one for each area. Areas from left and right hemispheres were considered separately, since some areas exhibit important asymmetries in their functional connectivity and/or spatial relationships with adjacent areas [9].

Fully-connected framework: An optimised FCNN architecture, which better reflects current practices, was used as a baseline to ensure a fair comparison with the gDL framework. Our optimised FCNN architecture differed from the original HCP FCNN [9] (Supplementary Methods: Section 6) in the following ways: 1) the number of hidden layers was increased from one ($H_1 = 9$) to three, with dimensions of 128 (H_1), 64 (H_2) and 32 (H_3); 2) tahn non-linearities were replaced with SiLU non-linearities; 3) batchnorm was used after each non-linearity; 4) the mean squared error loss was replaced with a binary cross-entropy with logits loss; 5) each model was trained for a minimum/maximum of 30/80 epochs. If validation area under the receiver-operator curve (AUROC) did not increase for 5 consecutive epochs, training was terminated and the model with the highest AUROC was selected.

Geometric deep learning framework: To address outstanding issues in generalising individualised HCP_MMP1.0 parcellations to the lower quality UKB imaging data, a novel geometric deep learning framework was implemented on icospheric cortical surfaces [8]. Each segmentation network (one per cortical area) was based on a U-Net [23] with 5 encoding and 5 decoding blocks (Figure 2). Each block contained a single MoNet convolutional layer parameterised with K = 37 Gaussian kernels and polar pseudo-coordinates [19], followed by a SiLU nonlinearity [6]. The optimal number of Gaussian kernels was determined by first assessing model performance for selected cortical areas using K = [5, 10, 15, 20, 25, 30, 35, 40, 45, 50] (optimal model performance when K = 35) and then further optimising by making unit adjustments of K. Downsampling and upsampling were achieved following the procedure described in [8]. Model overfitting was addressed by using 2D dropout ($p=0.2$) [24] prior to the first two downsamples. Training was carried out in two stages. During the first ('pretraining') stage, a reconstruction autoencoder was trained for the purpose of initialising the weights of the segmentation model for each cortical area. The reconstruction autoencoder had the same architecture as Figure 2 but lacked

[7] These are the same features as for [9] with the exclusion of task MRI features (1 mean task fMRI activation map, and 20 task fMRI spatial ICA maps) which had no correlates in UKB
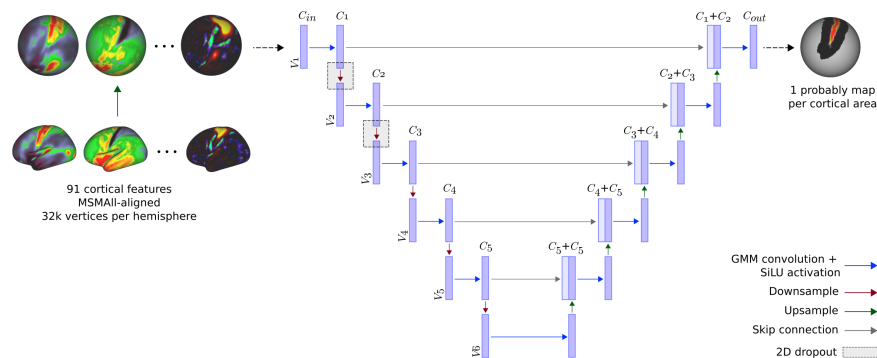
Fig. 2. Geometric deep learning architecture. C represents the number of channels, and V represents the number of vertices. For the segmentation networks: $[C_{in}, C_1, C_2, C_3, C_4, C_5, C_{out}] = [91, 32, 64, 128, 256, 512, 1]$. For the reconstruction autoencoder, skip connections were removed and C_{out} varied across cortical areas.

skip connections. Each reconstruction autoencoder was tasked with reconstructing only the most informative cortical features for a given area, as determined through visual inspection. All 91 cortical features were used as inputs in order to match the input weight dimensions for the segmentation model in the second training stage. However, the number of output channels (labels) to reconstruct was variable across cortical areas, depending on how informative each cortical feature was. The reconstruction autoencoder model weights were initialised with xavier initialisation [12]. Segmentation was performed in the second training stage, using the pretrained model weights. Models inputs were the same 91 cortical features, whilst labels were subject-specific parcellations derived from [9], generated using the original FCNN architecture. Each segmentation model output was transformed into a probability map using a sigmoid function, where values represented the probability that each vertex belongs to a given cortical area. Final parcellations were generated by aggregating across all classifiers using a winner-takes-all approach, with each vertex being assigned the label with the highest probability.

Implementation and training Models were trained and tested using the HCP210P and HCP29T subsets described in [9]. Cortical features and labels were upsampled from a standard resolution (32,492 vertices) to a sixth order icospheric resolution (40,962 vertices). All cortical features were normalised within-subject and within-feature to a mean of 0 and standard deviation of 1, and extreme values were clipped at $\pm$ 4 standard deviations. Autoencoder models were trained and validated using the mean absolute error, whilst segmentation models were trained using a binary cross-entropy loss with logits, and validated using area under the receiver-operator curve. All models were trained with a batch
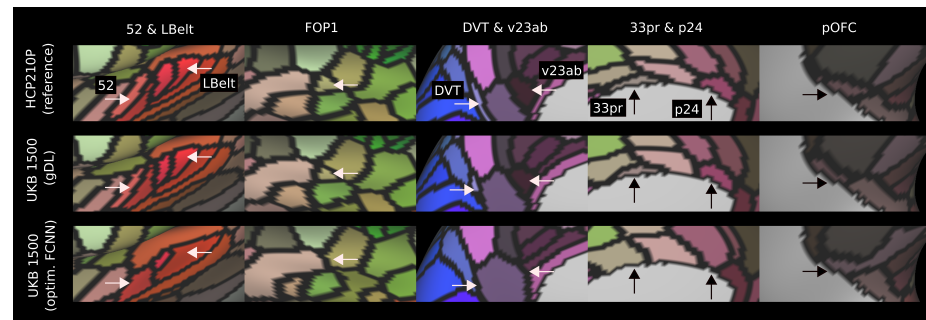
Fig. 3. Comparison of selected cortical areas across gDL and optimised FCNN architectures in UKB. All cortical areas highlighted are absent from the UKB 1500 group average parcellation when using optimised FCNN architecture. However, all cortical areas are present in UKB 1500 group average parcellation using gDL, and share very similar topography to HCP210P group average. Only left hemisphere shown for visualisation purposes.

size of 1, and optimisation was perfomed using AdamW [17] with a learning rate of 10^{-3}. All models were trained on a single NVIDIA RTX 24GB GPU, for a minimum/maximum of 15/30 epochs. Model training was terminated if there were no improvements in validation performance after 5 consecutive epochs. Once trained, segmentation models were then applied to all UKB subjects. To account for differences in the intensity distributions across each population, all UKB subject features were histogram matched to the group average of the HCP 210P feature maps used for training. No UKB subjects were seen during training.

3   Results & Discussion

Figure 3 compares the group average parcellation of 1500 UKB subjects using the optimised FCNN and gDL frameworks. Only 345 (left hemisphere: 172; right hemisphere: 173) of 360 cortical areas were present in the optimised FCNN group average, whilst all cortical areas were present in the results from our proposed gDL method. Of note, the original HCP FCNN framework [9] was unable to generate any meaningful parcellations for individual UKB data, most likely due to a combination of lower data quality in the UKB and the spatial independence of the FCNN. Moreover, using a single gDL network to parcellate all areas simultaneously resulted in a group average parcellation containing only 135 of 180 areas per hemisphere.

The proposed gDL model also outperformed the optimised FCNN architecture in average areal detection rate (99.1% vs. 88.7%) and lowest areal detection rate (80.9% vs. 29.9%) across both hemispheres. Importantly, these gDL results were not a result of each segmentation model predicting one consistent parcellation across subjects: UKB areal probability distributions were highly similar to those in the HCP, and demonstrate that cortical areas vary in location, even

after areal feature-based cortical alignment (Figure 4a, third row). Moreover, visual inspection indicated that our method was sensitive to atypical cortical topographies, with 3.7% (n = 54) of subjects having split 55b and 1.7% (n=26) having shifted 55b (Figure 4b).

Both the optimised FCNN and gDL frameworks performed well in unseen HCP210V cortical data, and demonstrate improvements compared to the original HCP FCNN used in [9] (Table 1). The smaller differences in performance between the optimised FCNN and gDL frameworks are expected, given that the amount and quality of MRI data per subject in the HCP is substantially higher compared to the UKB.

Despite marked differences in imaging protocols, these results highlight the excellent replicability of the HCP_MMP1.0 parcellation using the proposed gDL framework (Figure 4a), which is pertinent given concerns about reproducibility in neuroimaging [20], and stands in contrast to the replicability of other fully data-driven cortical parcellations in independent datasets [13,16]. However, an important limitation is that both optimised FCNN and gDL frameworks were trained using labels generated from the original HCP FCNN [9], which should not be considered ground truth. Further work is required to generate more accurate training labels, and to understand how these impact model performance and parcellation accuracy.

4   Conclusion

Here we present and evaluate an approach for generating individualised multi-modal cortical parcellations that generalise to markedly different datasets, and highlights the reproducibility of the HCP_MMP1.0 parcellation in independent cohorts.
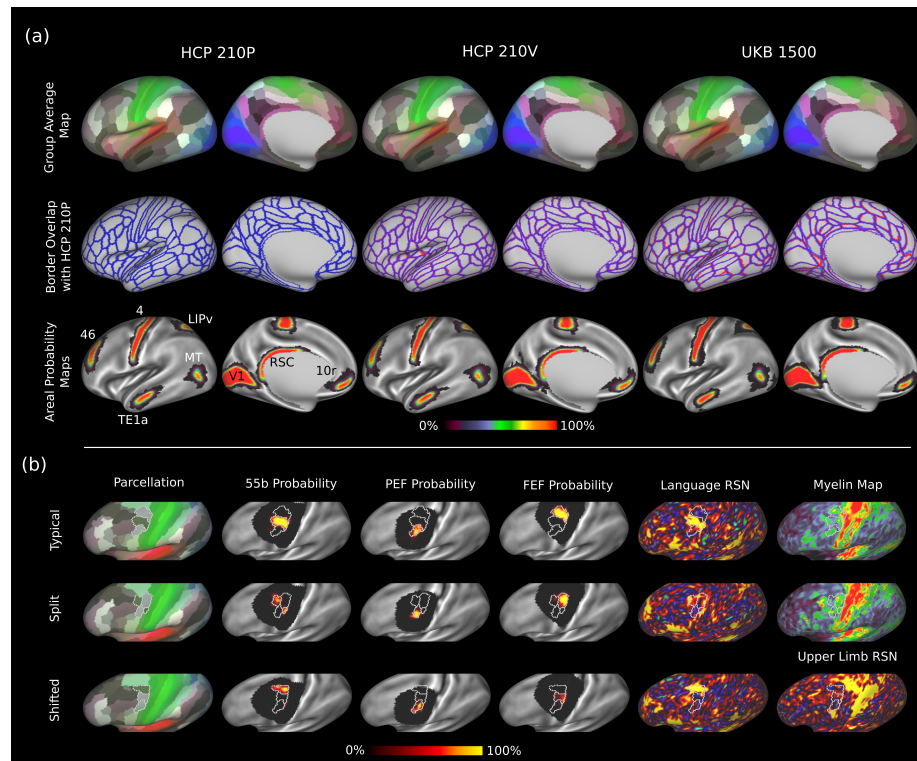
5   Acknowledgements

Model	Original FCNN [9]	Optimised FCNN	gDL
Average areal detection rate (%)	96.4	99.3	**99.9**
Lowest areal detection rate (%)	70.3	89.6	**96.4**
Split 55b (%)	4.6	**5.1**	1.7
Shifted 55b (%)	4.4	**5.6**	2.9

Table 1. Comparison of model performance in HCP 210V validation set. FCNN: fully-connected neural network; gDL: geometric deep learning

Fig. 4. Multimodal parcellation in the HCP and UKB. (a) First row: group average maps for HCP 210P, HCP 210V and 3000 UKB subjects. Second row: Border overlap between HCP 210P (blue; seen during training), HCP 210V and UKB 3000 (red; not seen during training). Areas of purple indicate overlap between borders. The Dice overlap ratio between HCP 210P and UKB 3000 was 0.986, relative to 0.996 between HCP 210P and 210V group maps. Third row: Probabilistic maps of areas V1, 4, RSC, MT, LIPv, TE1a, 46, and 10r, overlaid on MSMAll-aligned S1200 HCP average curvature map. As expected, spatial variability of cortical areas between subjects changes across the cortex, and this variability is consistent between HCP 210P and HCP 210V reported in [9] using a fully connected neural network, and UKB 3000 using our method. (b) Representative examples of typical (first row), split (second row) and shifted (third row) area 55b topographies. Alongside each individual parcellation are the probability maps for areas 55b, frontal eye field (FEF) and premotor eyefield (PEF), and the language spatial ICA map (from ICA d=40). T1w/T2w maps are shown for typical and split 55b topographies, whilst the upper limb spatial ICA map is shown for the shifted 55b topography [9]. All maps are displayed on an inflated cortical surface. Results were similar for both hemispheres, but only the left hemisphere is shown.

References

1. Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L., Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., et al.: Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. Neuroimage **166**, 400–424 (2018)
2. Assem, M., Glasser, M.F., Van Essen, D.C., Duncan, J.: A domain-general cognitive core defined in multimodally parcellated human cortex. Cerebral Cortex **30**(8), 4361–4380 (2020)
3. Bijsterbosch, J., Harrison, S.J., Jbabdi, S., Woolrich, M., Beckmann, C., Smith, S., Duff, E.P.: Challenges and future directions for representations of functional brain organization. Nature neuroscience **23**(12), 1484–1495 (2020)
4. Bijsterbosch, J.D., Woolrich, M.W., Glasser, M.F., Robinson, E.C., Beckmann, C.F., Van Essen, D.C., Harrison, S.J., Smith, S.M.: The relationship between spatial configuration and functional connectivity of brain regions. elife **7**, e32992 (2018)
5. Coalson, T.S., Van Essen, D.C., Glasser, M.F.: The impact of traditional neuroimaging methods on the spatial localization of cortical areas. Proceedings of the National Academy of Sciences **115**(27), E6356–E6365 (2018)
6. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks **107**, 3–11 (2018)
7. Elliott, L.T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K.L., Douaud, G., Marchini, J., Smith, S.M.: Genome-wide association studies of brain imaging phenotypes in uk biobank. Nature **562**(7726), 210–216 (2018)
8. Fawaz, A., Williams, L.Z., Alansary, A., Bass, C., Gopinath, K., da Silva, M., Dahan, S., Adamson, C., Alexander, B., Thompson, D., et al.: Benchmarking geometric deep learning for cortical segmentation and neurodevelopmental phenotype prediction. bioRxiv pp. 2021–12 (2021)
9. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al.: A multimodal parcellation of human cerebral cortex. Nature **536**(7615), 171–178 (2016)
10. Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L., Auerbach, E.J., Behrens, T.E., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., et al.: The human connectome project's neuroimaging approach. Nature neuroscience **19**(9), 1175–1187 (2016)
11. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
12. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
13. Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E.: Generation and evaluation of a cortical area parcellation from resting-state correlations. Cerebral cortex **26**(1), 288–303 (2016)
14. Krubitzer, L.: The magnificent compromise: cortical field evolution in mammals. Neuron **56**(2), 201–208 (2007)
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)

16. Lewis, J.D., Bezgin, G., Fonov, V.S., Collins, D.L., Evans, A.C.: A sub+ cortical fmri-based surface parcellation. Human Brain Mapping **43**(2), 616–632 (2022)

17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

18. Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al.: Multimodal population brain imaging in the uk biobank prospective epidemiological study. Nature neuroscience **19**(11), 1523–1536 (2016)

19. Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., Bronstein, M.M.: Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5115–5124 (2017)

20. Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.B., Vul, E., Yarkoni, T.: Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature reviews neuroscience **18**(2), 115–126 (2017)

21. Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., et al.: Multimodal surface matching with higher-order smoothness constraints. Neuroimage **167**, 453–465 (2018)

22. Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M.: Msm: a new flexible framework for multimodal surface matching. Neuroimage **100**, 414–426 (2014)

23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)

24. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)

25. Sydnor, V.J., Larsen, B., Bassett, D.S., Alexander-Bloch, A., Fair, D.A., Liston, C., Mackey, A.P., Milham, M.P., Pines, A., Roalf, D.R., et al.: Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. Neuron **109**(18), 2820–2846 (2021)

26. Wang, C., Martins-Bach, A.B., Alfaro-Almagro, F., Douaud, G., Klein, J.C., Llera, A., Fiscone, C., Bowtell, R., Elliott, L.T., Smith, S.M., et al.: Phenotypic and genetic associations of quantitative magnetic susceptibility in uk biobank brain imaging. Nature neuroscience **25**(6), 818–831 (2022)