


CSVR: Combined Surface and Volume Registration for brain MRI

Saga N.B. Masui^{1,2} 

SAGA.MASUI@KCL.AC.UK

Yourong Guo^{1,2}

YOURONG.GUO@KCL.AC.UK

Mohamed A. Suliman¹

MOHAMED.SULIMAN@KCL.AC.UK

Mattias P. Heinrich³

MATTIAS.HEINRICH@UNI-LUEBECK.DE

Nashira Baena²

NASHIRA.BAENA@KCL.AC.UK

Irina Grigorescu²

IRINA.GRIGORESCU@KCL.AC.UK

Logan Z. J. Williams^{1,2}

LOGAN.WILLIAMS@KCL.AC.UK

Ashleigh Davies^{1,6}

ASHLEIGH.DAVIES@KCL.AC.UK

Vanessa Kyriakopoulou^{1,2}

VANESSA.KYRIAKOPOULOU@KCL.AC.UK

Gráinne McAlonan⁴

GRAINNE.MCALONAN@KCL.AC.UK

Jonathan O’Muircheartaigh^{2,4,5}

JONATHANOM@KCL.AC.UK

Emma C. Robinson^{1,2,5,6}

EMMA.ROBINSON@KCL.AC.UK

¹ *Research Department of Biomedical Computing, School of Biomedical Engineering and Imaging Sciences, King’s College London, London, SE1 7EH, UK*

² *Centre for the Developing Brain, School of Biomedical Engineering and Imaging Science, King’s College London, London, SE1 7EH, UK*

³ *Institute of Medical Informatics, Universität zu Lübeck, Germany*

⁴ *Forensic and Neurodevelopmental Sciences, King’s College London, London, SE5 8AF, UK*

⁵ *MRC Centre for Neurodevelopmental Disorders, King’s College London, London, SE1 1UL, UK*

⁶ *Research Department of Early Life Imaging, School of Biomedical Engineering and Imaging Sciences, King’s College London, London, SE1 7EH, UK*

Editors: Under Review for MIDL 2026

Abstract

Nonlinear image registration is a cornerstone of neuroimaging analysis, supporting both qualitative and quantitative comparisons of brain structures across individuals and over time. While traditional volumetric registration methods, driven by voxel intensities, achieve good alignment of subcortical regions, they generally fail to capture correspondences between highly convoluted and variable cortical shapes. Surface-based methods, which instead regularise mappings as geodesics along the cortical sheet, yield improved cortical alignment but ignore the subcortical domain, limiting their utility for whole-brain analyses. A unified registration framework would address these limitations to enable integrated analysis of cortical and subcortical structures and the neuronal fibres that connect them. However, achieving this is challenging, since matching heterogeneous cortical shapes implies large volumetric displacements local to the cortex. To overcome these challenges, we introduce CSVR, the first deep learning-based framework for combined surface–volume registration. By integrating hierarchical registration strategies with discrete optimisation, CSVR achieves highly accurate, smooth, and anatomically plausible alignment of the entire brain.

Keywords: Image registration, Surface-based cortical registration, Discrete optimisation, Volumetric registration, Neuroimaging,

1. Introduction

Accurate spatial alignment of medical images is essential for both cross-sectional comparison of anatomical structures and longitudinal tracking of morphological changes, making image registration a foundational component of medical imaging pipelines. Yet for neuroimaging, the heterogeneity of human cortical morphology contradicts the core principles of classical image registration frameworks (Glasser et al., 2016; Amunts et al., 2007), which assume that all images can be diffeomorphically mapped to a common coordinate space where analogous anatomical structures overlap. In reality, cortical folds vary across individuals in terms of their number, branching and orientations (Guo et al., 2025; Thompson et al., 1996; Ono et al., 1990; Guillon et al., 2024) and since this variability is often entirely natural and unrelated to disease or cognition, it translates into experimental noise that confounds downstream analysis.

One way of addressing this problem has been to selectively compare individual brains with others that share similar patterns of cortical folding (Meng et al., 2016; Duan et al., 2017). Taking this logic further, there has also been a move toward hierarchical image registration frameworks (Guo et al., 2025; Ahmad et al., 2019; Dong et al., 2018) that incrementally align brains of increasingly dissimilar shapes by e.g. clustering individuals into groups that share common anatomies; then using these groups to generate a *family* of templates that represent key modes of shape variation; before gradually registering subjects through this hierarchy until all brains are brought into a single reference space (Guo et al., 2025).

Since matching and aligning complex shapes in 3D is a highly ill-posed problem, this clustering approach can be made more tractable through the use of spherical registration (Suliman et al., 2022, 2025; Robinson et al., 2014, 2018; Yeo et al., 2009), which simplifies the problem of cortical alignment to one of matching feature maps defined over smooth 2D spheres (Fischl et al., 1999; Drury et al., 1996). In this way, spherical image registration methods have repeatedly demonstrated their capacity for improving cortical shape and/or cytoarchitectonic correspondence, relative to volumetric approaches (Glasser et al., 2016; Coalson et al., 2018), to support more precise characterisation of cortical asymmetries (Williams et al., 2023; Meyer et al., 2014; Raznahan et al., 2011), neurodevelopmental growth (Garcia et al., 2018; da Silva et al., 2025), and functional organisation (Glasser et al., 2016; Yeo et al., 2011; Kong et al., 2019).

However, while spherical mapping approaches may improve cortical alignment, they come at the cost of losing geometric correspondence to the rest of the brain, meaning that surface and volume analyses are typically run in disconnected template spaces that are not aligned and cannot be directly compared (Wu et al., 2018). These challenges disproportionately impact structural connectivity studies which seed from the cortex but project through white matter volumes.

A holistic framework for combined surface and volume mapping could address these limitations. Yet, achieving this is non-trivial due to the challenge of diffusing spherical warps—that allow sulci and gyri to overlap—into the volume without inducing extreme distortions. Previous solutions have attempted to regularise this task through slow iterative processes (Postelnicu et al., 2008) with prohibitively long (~ 17 -hour) run-times, or limit

processing to within-subject longitudinal alignment (Gibson et al., 2009) or alignment of twins (that share similar brain shapes)(Lepore et al., 2010).

One increasingly cited solution for non-convex image matching problems has been to use discrete optimisation, which reframes image matching as a combinatorial optimisation problem, in which the optimal displacement field must be selected from a quantised set of possible options. By constraining deformations in this way, these algorithms are able to reject locally appealing but globally inconsistent configurations, to model large and highly deformable transformations while maintaining anatomically plausible warps (Robinson et al., 2014, 2018; Besenczi et al., 2024; Glocker et al., 2008). Recent advances in deep discrete registration (e.g., DDR (Suliman et al., 2022); GeoMorph (Suliman et al., 2025); PDD (Heinrich, 2019)) take this one step further to combine the geometric robustness of such discrete formulations with the efficiency of deep feature extractors, making them particularly well-suited to coupled cortical and volumetric alignment.

Contributions: This paper presents a holistic framework for cluster-based deep-discrete surface and volumetric registration (CSV) for precise anatomical alignment of individual human brains. It integrates deep discrete registration (DDR) (Suliman et al., 2022) of cortical surfaces with probabilistic dense displacement (PDD) (Heinrich, 2019) of MRI volumes within a single jointly-optimised framework to improve the alignment of whole-brain MRI more precisely than either surface or volume registration alone. Results show the resulting warps are smooth and biologically plausible, and generalise robustly to unseen examples. In particular, CSV performs well on all subjects, even more heterogeneous morphologies, where traditional volumetric methods show limited success.

2. Methods

Let $M = \{\mathbf{x}^M\}$ and $T = \{\mathbf{x}^T\}$ denote affinely aligned moving and target brain volumes, with $\mathbf{x}^M, \mathbf{x}^T \in \mathbb{Z}^3$ representing voxel coordinates. Let corresponding cortical surfaces be represented by mesh pairs $\partial M = (\mathcal{A}_M, \mathcal{S}_M^2)$ and $\partial T = (\mathcal{A}_T, \mathcal{S}_T^2)$ of corresponding anatomical and spherical surfaces that share vertex correspondences. In each case, anatomical surfaces $\mathcal{A}_M = \{\mathbf{v}_i^M \in \mathbb{R}^3\}_{i=1}^{N_v}$ and $\mathcal{A}_T = \{\mathbf{v}_i^T \in \mathbb{R}^3\}_{i=1}^{N_v}$ correspond to the inner cortical, or white matter, boundary defined from their corresponding volumes, from which spheres $\mathcal{S}_M^2 = \{\mathbf{s}_i^M \in \mathbb{R}^3\}_{i=1}^{N_v}$ and $\mathcal{S}_T^2 = \{\mathbf{s}_i^T \in \mathbb{R}^3\}_{i=1}^{N_v}$ are generated through learning-based spherical projection (Ma et al., 2025).

2.1. Deep Discrete Alignment

Deep discrete alignment frames image registration as a classification problem in which displacements – for a low resolution control point (CP) grid - are learnt through offering each (CP) $\{c_i\}_{i=1}^{N_c}$ a choice of target locations $\{l_i\}_{i=1}^{N_l}$ to deform to (Figure 2.1). For DDR, the control point grid and label points are defined from vertices derived from regularly tessellated icospheres. For PDD, the control point grid and label points are defined in 3D. Unlike classical discrete optimisation frameworks (Glocker et al., 2008; Robinson et al., 2014, 2018; Heinrich et al., 2015), the label space is made as wide as computationally possible such that the final deformation field may be classified in one shot. Regularisation is imposed through deep conditional random fields (CRFs) that enforce spatial smoothness by encouraging neighbouring control points to deform to similar target locations (Zheng et al., 2015).

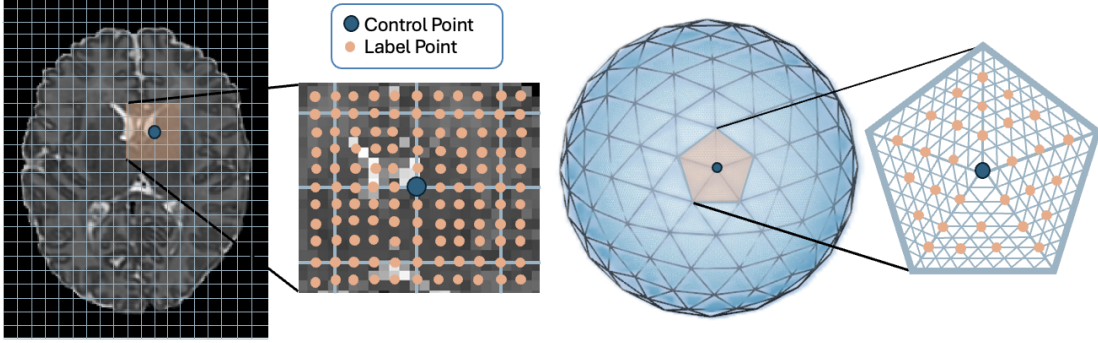


Figure 2.1: Control grids and label spaces of discrete registration in the spherical (top) and volumetric (bottom) domains. Modified from Robinson et al. (2014)

2.2. Proposed Framework

Our overall goal is to estimate a mapping such that both internal intensities or MRI volumes and cortical surface features (and geometry) are aligned. CVSR (Figure 2.2) achieves this via:

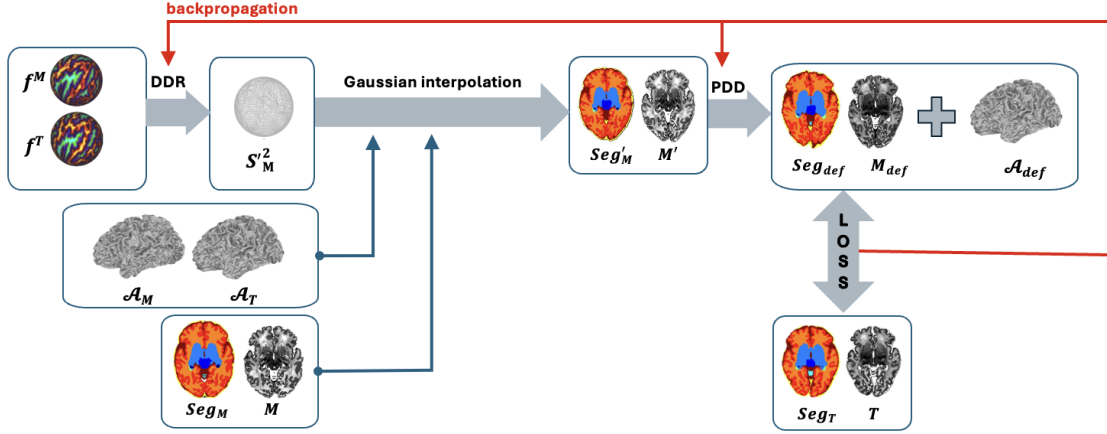


Figure 2.2: The CSVR framework: Moving and target sulcal depth maps are projected onto a regular icosphere and input to DDR to produce a deformed input sphere S_M^2 . This spherical warp is used to project the moving anatomical mesh \mathcal{A}_M onto the target anatomy \mathcal{A}_T to return a deformed anatomy \mathcal{A}'_M (with the shape of the target but topology of the input). Subtracting \mathcal{A}_M and \mathcal{A}'_M yields a sparse 3D displacement field $\{\mathbf{d}_i\}_{i=1}^{N_v}$, which is converted into a dense warp \mathbf{u}_s through differentiable Gaussian interpolation, then inverted pull back intensities from the moving volume onto the target grid (as per convention) to generate a deformed and resampled input volume M' , which is used as inputs for PDD. This process is jointly optimised to learn a combined surface-and-volume transformation.

2.2.1. SURFACE-DRIVEN ALIGNMENT (DDR)

Spherical alignment is performed using a pretrained DDR network from (Guo et al., 2025), that learns a spatial transformation $\phi_s : \mathcal{S}_M^2 \rightarrow \mathcal{S}_T^2$ that optimises overlap of sulcal depth functions (f^M, f^T) over three stages: (1) a rotation network $h_R(f^M, f^T)$ first estimates a 3D rotation matrix $\mathbf{R}_S \in \mathbb{R}^{3 \times 3}$ that corrects the residual misalignment between spheres that remains after volumetric affine initialisation; (2) a non-rigid deformation module $h_{NR}(f^M \circ \mathbf{R}_S \mathcal{S}_M^2, f^T)$ is then trained to classify the target location of each control point $\{c_i^{DDR}\}_{i=1}^{N_c} \in \mathcal{S}_C^2$ by choosing from a fixed set of labels $\{l_i^{DDR}\}_{i=1}^{N_l} \in \mathcal{S}_L^2$, where the deformation grid \mathcal{S}_C^2 and the label grid \mathcal{S}_L^2 correspond to regularly tessellated icospheres of different resolutions ($|\mathcal{S}_L^2| > |\mathcal{S}_C^2|$); (3) this process is smoothed through implementing CRF regularisation through a recurrent neural network (RNN) (Zheng et al., 2015), combined with a diffusion regulariser that operates on the spatial gradients. Both rotation and deformation modules employ spherical U-Net architectures (Ronneberger et al., 2015) that take sulcal depth features as input (resampled to a sixth-order isosurface) and learn MoNet convolutional filters (Monti et al., 2017). The final warp (output from h_{NR}) corresponds to a deformed spherical configuration $\mathcal{S}_M'^2$. To transfer this to the anatomical domain, we exploit vertex correspondence between \mathcal{S}_M^2 and \mathcal{A}_M (and \mathcal{S}_T^2 and \mathcal{A}_T), and implement differentiable barycentric interpolation \mathcal{I}_B . Using the step-by-step framework from (Robinson et al., 2018) this allows moving spherical vertices \mathbf{u}'^M_i to be located relative to target vertices \mathbf{u}^t_i , from which barycentric weights may be derived to project the moving anatomical mesh onto the target anatomical mesh. This deformed surface \mathcal{A}'_M is then used to extract a sparse displacement field $\{\mathbf{d}_i = \mathbf{v}'^M_i - \mathbf{v}^M_i\}_{i=1}^{N_v}$.

2.2.2. SURFACE-TO-VOLUME INTERPOLATION

From this sparse vertex-wise displacement we require a dense volumetric displacement field $\mathbf{u}_s(\mathbf{x})$, which is derived from a two-stage, differentiable Gaussian interpolation scheme. For each voxel location \mathbf{x}^M , we compute a weighted combination of displacements from nearby cortical vertices, such that:

$$\mathbf{u}_s^{\text{sparse}}(\mathbf{x}_j^M) = \sum_{i \in \mathcal{N}(\mathbf{x}^M)} w_{i,j} \mathbf{d}_i, \quad \text{with} \quad w_{i,j} = \exp \left(-\frac{\|\mathbf{v}_i^M - \mathbf{x}_j^M\|^2}{2\sigma_{\text{scatter}}^2} \right) \quad (1)$$

Here, $\mathcal{N}(\mathbf{x}^M)$ defines a cubic neighbourhood surrounding \mathbf{x}^M . This sparse field is then diffused into subcortical regions via a separable 3D Gaussian convolution with standard deviation σ_{vox} :

$$\mathbf{u}_s(\mathbf{x}) = G_{\sigma_{\text{vox}}} * \mathbf{u}_s^{\text{sparse}}(\mathbf{x}) \quad (2)$$

The final displacement field is normalised to preserve displacement magnitudes.

Finally, because volumetric registration networks typically predict *backward* deformation fields — i.e., mappings $\phi_v : T \rightarrow M$ that sample intensities as $M(\mathbf{x}^T + \mathbf{u}_v(\mathbf{x}^T))$, where voxels $\mathbf{x} \in T$ — this surface-derived field is first inverted before being used to transform M and its corresponding segmentation onto the reference configuration, yielding preliminary warped volumes (M') that propagate cortical correspondences between \mathcal{A}_M and \mathcal{A}'_M into the voxel domain.

2.2.3. VOLUMETRIC SUBCORTICAL ALIGNMENT (PDD):

M' is then passed to a pretrained probabilistic dense displacement network driven by image intensities I^M , I^T , and weakly supervised by tissue segmentation maps Seg^M and Seg^T (Heinrich, 2019). The model comprises three components: (1) an OBELISK-based feature extractor that produces intensity-invariant descriptors for M' and T (Heinrich et al., 2019), preceded by a $5 \times 5 \times 5$ convolution that captures edge-like information; (2) a correlation layer that considers possible displacement vectors and evaluates feature dissimilarity from negated MSE across all candidates; and (3) architecturally embedded regularisation that alternates approximate min-convolutions and mean-field inference to simulate a Markov Random Field (Krähenbühl and Koltun, 2011; Zheng et al., 2015) and enforce smoothness. Ultimately, PDD uses the regularised MSE feature dissimilarities to predict a probability distribution over displacement labels at each control point via a softmax layer, which drives a non-local (MSE-based) segmentation loss. A continuous displacement field is then obtained through probability-weighted averaging and upsampling and subjected to a deterministic diffusion regularisation penalty. Since PDD was originally developed for abdominal CT registration, we adapt this method to subcortical brain MRI by complementing these terms with deterministic Dice and mutual information losses evaluated on the final warped image, yielding a total loss of:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{nonlocal}}(\text{Seg}_{\text{prob}}^{\text{LR}}) + \mathcal{L}_{\text{MI}}(\text{I}_{\text{det}}^{\text{LR}}) + \mathcal{L}_{\text{Dice}}(\text{Seg}_{\text{det}}^{\text{LR}}) + \mathcal{L}_{\text{reg.diffusion}}(\text{I}_{\text{det}}^{\text{HR}}) \quad (3)$$

Here, all losses are calculated at low resolution (LR), on either probabilistic or deterministic images and labels, but the diffusion regularisation is applied high resolution (HR) and the approximated MRF implicitly regularises final deformation via the probabilistic weights of the candidate displacements. We also allow non-isotropic, dataset-adaptive control-grid dimensions and introduce a cortical deformation penalty to minimise interference with $\mathbf{u}_s(\mathbf{x})$.

2.2.4. JOINT OPTIMISATION AND WARP COMPOSITION:

After pretraining DDR and PDD to separately optimise cortical and volumetric alignment, we then refine the models through differentiable Gaussian Interpolation (Section 2.2.2) into a unified model, CSVr, that jointly optimises both networks by backpropagating the PDD-losses, complemented with an additional high-resolution Dice loss, through both networks.

3. Experimental Methods and Implementation

3.1. Data and Processing

A total of 681 subjects from the developing Human Connectome Project (dHCP) and 100 subjects from Brain Imaging in Babies (BIBS) were used for the training and evaluation of our models, respectively (Edwards et al., 2022). Out of these neonates 648 were term-born (scan age 41.51.7 weeks, 297 biologically female) and 133 were preterm-born (scan age: 41.31.7 weeks, 59 biologically female). Both datasets were acquired at the Evelina Newborn Imaging Centre using a Philips Achieva 3T scanner with a dedicated 32-channel neonatal head coil and positioning device (Glasser et al., 2013; Hughes et al., 2017). T2-weighted

images were acquired in two stacks (sagittal/axial), with a repetition time (TR)/echo time (TE) of 12 000/156ms, flip angle of 90°, in-plane resolution of 0.8 x 0.8mm, slice thickness of 1.6 and overlap of 0.8mm. These were motion-corrected and super-resolved to generate output volumes with 0.5 mm isotropic grid sizes (Cordero-Grande et al., 2016, 2018). All babies were imaged during natural sleep.

T2-weighted images were histogram-normalised and preprocessed using the dHCP Deep Learning-based Neonatal Pipeline (Ma et al., 2025), which builds on and is trained using outputs from the classical dHCP structural pipeline (Schuh et al., 2017). This affinely aligns volumes to the dHCP 40-week neonatal T2w template (Makropoulos et al., 2018); then uses learning-based mesh fitting to shrink a template surface to the inner cortical (white) boundary, which is then reinflated to meet the outer cortical (pial surface). White surfaces are subsequently inflated using a GPU-accelerated reimplement of FreeSurfer’s inflation algorithm (Fischl, 2012), from which sulcal depth maps were estimated by integrating over the functional. Finally, spheres are generated through spherical mapping, optimised using a spherical U-Net implementation similar to (Zhao et al., 2021). Prior to registration using CSVr, sulcal depth maps and white matter surfaces were resampled to ico-6 resolution, all volumetric images were histogram normalised and resampled to dimensions of $H = 176$, $W = 224$, $D = 160$.

3.2. Clustering

To simplify the registration problem and minimise extreme distortions, all subjects were first clustered into groups that share more similar folding patterns using the method outlined in (Guo et al., 2025). In total, 90 clusters were used (30 for each of the frontal, parietal and temporal lobes) due to observations that cortical folding variants do not co-occur, i.e. individuals that share more similar folding patterns of the frontal lobes do not necessarily share variants of parietal or temporal lobes. To generate these clusters, all subjects underwent initial pairwise surface registration, using DDR (Suliman et al., 2022), which evaluates the similarity of their folding through assessing their overlap following diffeomorphic alignment. This means that the training and testing datasets consist of pairs of individuals with broadly similar folding patterns (for any of their lobes). Importantly, to avoid data leakage, all samples from each cluster were first separated into training (dHCP) and test (BIBS) subsets before pairing was performed, and each pair is applied twice - with each sample used once as moving and once as the target. Using the standard convention, the dataloader randomises the order in which the pairs are presented during training.

3.3. Training and Model Implementation

DDR (Surface Registration). The original DDR paper (Suliman et al., 2022) trained subject to template alignment. Therefore, here, we utilise the modified *pairwise* training scheme of (Guo et al., 2025). Data is input at high resolution - resampled to a sixth-order icosphere (ico-6) with 40,962 vertices. Labels are also defined at ico-6 resolution, and set to the 80 nearest neighbours of each control point, defined on an ico-2 grid (162 vertices). Following CRF optimisation, the deformed low resolution control grid is upsampled to the input data resolution (ico-6) using barycentric interpolation. Optimisation was performed

using Adam with a learning rate of 10^{-4} , a diffusion regularisation weight of $\lambda = 1$, and performance was assessed via five-fold cross-validation on held-out subject pairs.

PDD (Volumetric Registration). PDD was trained using: a diffusion regularisation weight $\lambda = 1$, dice weight of 0.5, MI weight of 1 and non-local label weight of 15. The latter was scaled to account for inherent differences in magnitudes between the losses. We used control grids of size $44 \times 56 \times 40$ (which corresponds to sampling the input volume by 4). The network was optimised using the Adam optimiser with a learning rate of 5×10^{-3} , for 5 epochs over approximately 126k pairs. Model checkpoints were evaluated every 2000 iterations on a small validation subset containing at least one subject pair from each anatomical cluster, and the best-performing checkpoint was retained. Notably, we found that accurate alignment was achieved after only 2h 45 min of training, and more specifically after the first 30k iterations. This aligns well with Heinrich (2019)’s original finding that PDD only requires a small number of subjects and a short training time.

CSVR. After independent pretraining, the combined model, CSVr, was trained on a subset of subjects representative of all anatomical clusters, but with permutations capped at 30 subjects per cluster, motivated by the promising generalisation performance of PDD. Training used the same hyperparameters as the pretrained networks, with the addition of a high-resolution Dice loss term (weight 0.5). Before producing a final output, the inverted surface warp was composed with the volumetric warp to produce a single deformation field $\mathbf{u}(\mathbf{x}) = \mathbf{u}_v \circ \mathbf{u}_s^{-1}$, applied to the original moving image as $M_{\text{registered}}(\mathbf{x}) = M(\mathbf{x} + \mathbf{u}(\mathbf{x}))$, and to the surface as $\mathcal{A}_{M,\text{registered}} = \{\mathbf{v}_i + \mathbf{u}^{-1}(\mathbf{v}_i)\}_{i=1}^{N_v}$. The purpose of this was to minimise interpolation artefacts.

4. Results and Discussion

The inference time of CSVr is only 0.73s, yielding anatomically plausible displacement fields with less than 0.5% negative Jacobians. We evaluate CSVr against three baseline volumetric methods: affine alignment (using ANTsPy (Avants et al., 2011)), FMRIB’s Nonlinear Image Registration Tool, FNIRT (Andersson et al., 2007), and PDD without the cortical penalty term (Figure 4.1). FNIRT was configured with a control grid of the same resolution as PDD, and our whole-brain PDD network was trained using the parameters specified above. All methods used identical preprocessing and initialisation procedures (Section 3.1). Registration accuracy was quantified using Dice overlap coefficients computed across DrawEM tissue labels. Statistical significance was assessed per tissue type using paired two-tailed t-tests with false discovery rate (FDR) correction for multiple comparisons.

CSVR achieved statistically significant improvements over all baseline methods across anatomical tissue types (all $p < 0.001$, FDR corrected) (1). For cortical gray matter specifically, CSVr demonstrated mean Dice improvements of +0.127 over affine alignment, +0.07 over FNIRT, and +0.038 over PDD. Across all tissue types, improvements versus affine ranged from +0.094 (deep gray matter) to +0.148 (ventricles), versus FNIRT from +0.027 (white matter) to +0.124 (hippocampi and amygdala), and versus PDD from +0.001 (deep gray matter, $p = 0.59$) to +0.031 (cortical gray matter). These results demonstrate that cortical surface constraints benefit not only cortical alignment but also propagate improvements to subcortical structures.

Region	Affine	FNIRT	PDD	CSV
Cortical Gray Matter	0.5282	0.5801	0.5771	0.6150
Deep Gray Matter	0.8806	0.8897	0.9231	0.9239
Hippocampi and Amygdala	0.6813	0.7534	0.7800	0.8034
Ventricles	0.5287	0.6059	0.7141	0.7271
White Matter	0.6160	0.6534	0.6542	0.6809
Overall Mean	0.6470	0.6965	0.7297	0.7501

Table 1: Average Model Dice by Brain Region

While subjects were grouped into registration clusters based on cortical folding patterns (specifically sulcal depth similarity), these do not always translate to similar whole-brain morphology. Consequently, some subject pairs exhibit substantially different T2-weighted intensities and overall brain shape, presenting challenging registration scenarios where affine or conventional nonlinear methods struggle (Figure 4.2). Notably, CSV maintains robust performance even for these difficult cases, with much improved alignment, in particular relative to affine and FNIRT, with up to 40% increase in dice overlap for some of the most challenging subjects. This pattern suggests that PDD’s discrete optimisation framework confers robustness to local minima associated with highly heterogeneous brain morphology, while CSV’s cortical surface constraints provide additional gains primarily for cases where cortical folding patterns are informative for whole-brain alignment. Figure A.1 further highlights CSV’s ability to align difficult cases.

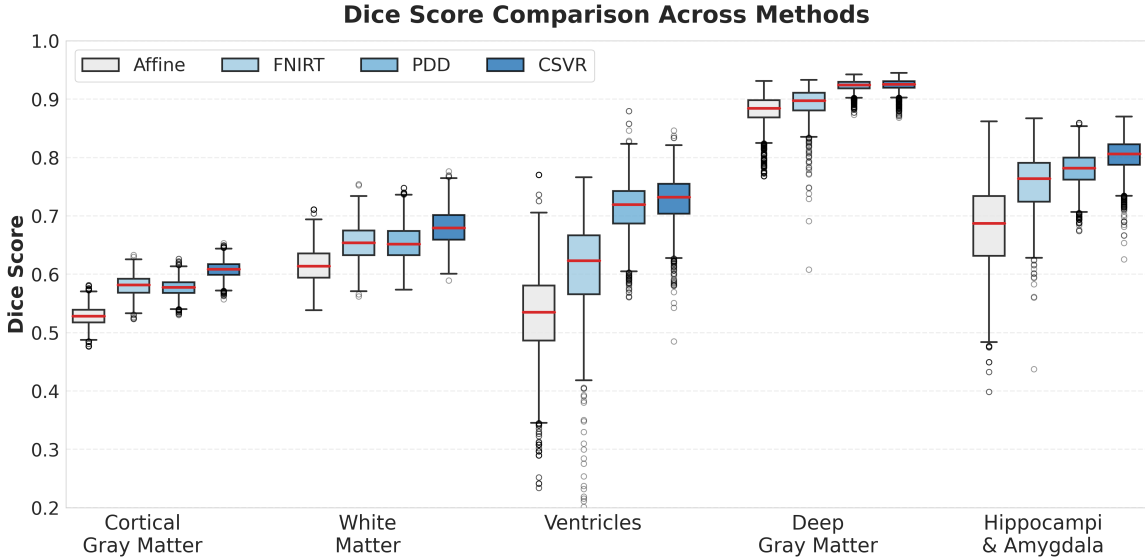


Figure 4.1: Dice Overlap between brain regions. CSV performs better than other methods across all tissue types, with the exception of deep gray matter, where alignment is on par with PDD. For comparison, the average overall Dice across corresponding unregistered brains was 0.3659.

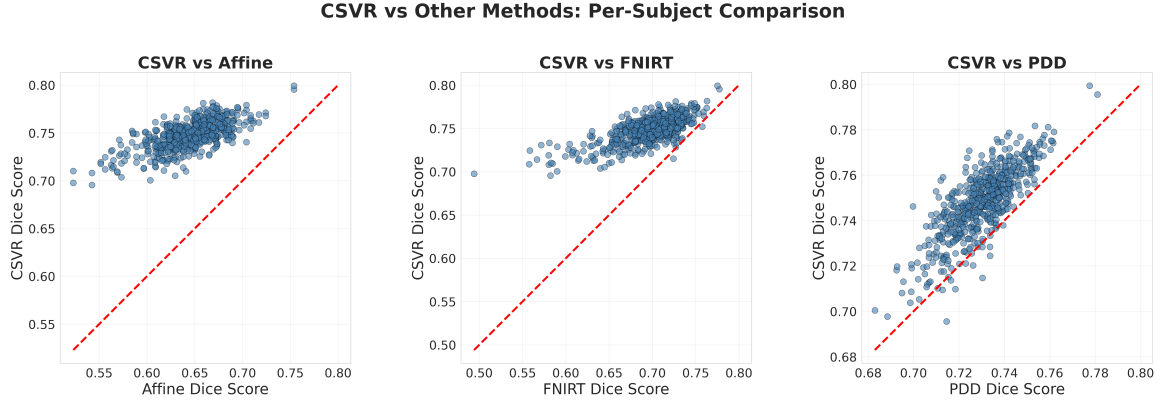


Figure 4.2: Per-Subject Comparison of CSVr vs other methods. Scatter plots comparing the average Dice score of CSVr against Affine, FNIRT, and PDD registration methods for individual subject pairs. The dashed red line represents the identity line ($y = x$), where no performance change occurs. Points above the line indicate superior performance by CSVr. CSVr demonstrates widespread improvement over Affine and FNIRT across the entire range of registration difficulty (Dice scores), suggesting robust performance even in challenging cases. The performance difference is reduced when comparing CSVr to PDD for the most difficult subject pairs (lower Dice scores), indicating PDD’s robustness to challenging whole-brain morphology.

The current implementation of CSVr operates on relatively low-resolution control point and label grids, which, while enabling rapid inference times, may limit the precision of fine-scale anatomical alignment. To achieve higher registration accuracy, future developments will include adopting a multi-stage optimisation strategy wherein surface and volumetric registrations are iteratively refined across progressively finer resolutions. Such a hierarchical approach would allow coarse alignment to establish global correspondence before fine-tuning local deformations, further improving Dice scores while maintaining computational efficiency.

5. Conclusion

In conclusion, we introduce CSVr, the first deep discrete combined surface-volume registration network for brain MRI. Our method represents a significant step toward addressing the longstanding challenge of combined surface-volumetric template construction; currently, whole-brain analyses are constrained by the lack of integrated templates, but by jointly optimising cortical surface and volumetric alignment, CSVr provides the foundational method necessary for generating such templates within [Guo et al. \(2025\)](#)’s hierarchical framework. While we have shown that CSVr is able to accurately register subjects within clusters of similar cortical folding, we have also shown that the model performs particularly well on more heterogeneous, difficult registration cases, showing great promise for downstream inter-cluster alignment and template creation.

Acknowledgments

The authors would like to acknowledge the participants of the dHCP and BIBS studies. The dHCP neonatal dataset was provided by the developing Human Connectome Project, KCL-Imperial-Oxford Consortium funded by the European Research Council (ERC) under the European Union Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. [319456]. The BIBS data was funded by EU-AIMS (European Autism Interventions)/EU AIMS-2-TRIALS, an Innovative Medicines Initiative Joint Undertaking under Grant Agreement No. 777394. The School of Biomedical Engineering and Imaging Sciences is supported by the Wellcome EPSRC Centre for Medical Engineering at King’s College London (WT 203148/Z/16/Z) and the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre (BRC) award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London and King’s College Hospital NHS Foundation Trust. We acknowledge infrastructure support from the NIHR Mental Health BRC at South London and Maudsley NHS Foundation Trust, King’s College London.

References

- Sahar Ahmad, Jingfan Fan, Pei Dong, Xiaohuan Cao, Pew-Thian Yap, and Dinggang Shen. Deep learning deformation initialization for rapid groupwise registration of inhomogeneous image populations. *Frontiers in Neuroinformatics*, 13:34, 2019.
- Katrin Amunts, Axel Schleicher, and Karl Zilles. Cytoarchitecture of the cerebral cortex—more than localization. *Neuroimage*, 37(4):1061–1065, 2007.
- JL Andersson, Mark Jenkinson, Stephen Smith, et al. Non-linear registration, aka spatial normalisation fmrib technical report tr07ja2. *FMRIB Analysis Group of the University of Oxford*, 2(1):e21, 2007.
- Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
- Renato Besenczi, Yourong Guo, and Emma C Robinson. High performance groupwise cortical surface registration with multimodal surface matching. In *International Workshop on Biomedical Image Registration*, pages 321–332. Springer, 2024.
- Timothy S Coalson, David C Van Essen, and Matthew F Glasser. 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.
- Lucilio Cordero-Grande, Rui Pedro AG Teixeira, Emer J Hughes, Jana Hutter, Anthony N Price, and Joseph V Hajnal. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Transactions on Computational Imaging*, 2(3):266–280, 2016.
- Lucilio Cordero-Grande, Emer J Hughes, Jana Hutter, Anthony N Price, and Joseph V Hajnal. Three-dimensional motion corrected sensitivity encoding reconstruction for multi-

- shot multi-slice mri: application to neonatal brain imaging. *Magnetic resonance in medicine*, 79(3):1365–1376, 2018.
- Mariana da Silva, Kaili Liang, Kara E Garcia, M Jorge Cardoso, and Emma C Robinson. Differential patterns of cortical expansion in fetal and preterm brain development. *bioRxiv*, 2025.
- Pei Dong, Xiaohuan Cao, Pew-Thian Yap, and Dinggang Shen. Efficient groupwise registration of mr brain images via hierarchical graph set shrinkage. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 819–826. Springer, 2018.
- Heather A Drury, David C Van Essen, Charles H Anderson, Chris W Lee, Thomas A Coogan, and James W Lewis. Computerized mappings of the cerebral cortex: a multiresolution flattening method and a surface-based coordinate system. *Journal of cognitive neuroscience*, 8(1):1–28, 1996.
- Dingna Duan, Shunren Xia, Yu Meng, Li Wang, Weili Lin, John H Gilmore, Dinggang Shen, and Gang Li. Exploring gyral patterns of infant cortical folding based on multi-view curvature information. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 12–20. Springer, 2017.
- A David Edwards, Daniel Rueckert, Stephen M Smith, Samy Abo Seada, Amir Alansary, Jennifer Almalbis, Joanna Allsop, Jesper Andersson, Tomoki Arichi, Sophie Arulkumaran, et al. The developing human connectome project neonatal data release. *Frontiers in neuroscience*, 16:886772, 2022.
- Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.
- Bruce Fischl, Martin I Sereno, Roger BH Tootell, and Anders M Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, 8(4):272–284, 1999.
- Kara E Garcia, Emma C Robinson, Dimitrios Alexopoulos, Donna L Dierker, Matthew F Glasser, Timothy S Coalson, Cynthia M Ortinau, Daniel Rueckert, Larry A Taber, David C Van Essen, et al. Dynamic patterns of cortical expansion during folding of the preterm human brain. *Proceedings of the National Academy of Sciences*, 115(12):3156–3161, 2018.
- Eli Gibson, Ali R Khan, and Mirza Faisal Beg. A combined surface and volumetric registration (savor) framework to study cortical biomarkers and volumetric imaging data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 713–720. Springer, 2009.
- Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

- Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.
- Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12(6):731–741, 2008.
- Louise Guillon, Joël Chavas, Audrey Bénézit, Marie-Laure Moutard, Pauline Roca, Charles Mellerio, Catherine Oppenheim, Denis Rivière, and Jean-François Mangin. Identification of rare cortical folding patterns using unsupervised deep learning. *Imaging Neuroscience*, 2:1–27, 2024.
- Yourong Guo, Mohamed A Suliman, Logan ZJ Williams, Renato Besenczi, Kaili Liang, Simon Dahan, Vanessa Kyriakopoulou, Grainne M McAlonan, Alexander Hammers, Jonathan O’Muircheartaigh, et al. Motifs of brain cortical folding from birth to adulthood: structural asymmetry and folding-functional links. *bioRxiv*, pages 2025–03, 2025.
- Mattias P Heinrich. Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 50–58. Springer, 2019.
- Mattias P Heinrich, Oskar Maier, and Heinz Handels. Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities. *Visceral Challenge@ ISBI*, 1390:27, 2015.
- Mattias P Heinrich, Ozan Oktay, and Nassim Bouteldja. Obelisk-net: Fewer layers to solve 3d multi-organ segmentation with sparse deformable convolutions. *Medical image analysis*, 54:1–9, 2019.
- Emer J Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N Price, et al. A dedicated neonatal brain imaging system. *Magnetic resonance in medicine*, 78(2):794–804, 2017.
- Ru Kong, Jingwei Li, Csaba Orban, Mert R Sabuncu, Hesheng Liu, Alexander Schaefer, Nanbo Sun, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, et al. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral cortex*, 29(6):2533–2551, 2019.
- Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011.
- Natasha Lepore, Anand A Joshi, Richard M Leahy, Caroline Brun, Yi-Yu Chou, Xavier Pennec, Agatha D Lee, Marina Barysheva, Greig I De Zubicaray, Margaret J Wright, et al. A new combined surface and volume registration. In *Medical Imaging 2010: Image Processing*, volume 7623, pages 448–456. SPIE, 2010.

- Qiang Ma, Kaili Liang, Liu Li, Saga Masui, Yourong Guo, Chiara Nosarti, Emma C Robinson, Bernhard Kainz, and Daniel Rueckert. The developing human connectome project: A fast deep learning-based pipeline for neonatal cortical surface reconstruction. *Medical Image Analysis*, 100:103394, 2025.
- Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.
- Yu Meng, Gang Li, Li Wang, Weili Lin, John H Gilmore, and Dinggang Shen. Discovering cortical folding patterns in neonatal cortical surfaces using large-scale dataset. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer, 2016.
- Martin Meyer, Franziskus Liem, Sarah Hirsiger, Lutz Jäncke, and Jürgen Hänggi. Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex. *Cerebral Cortex*, 24(10):2541–2552, 2014.
- Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5115–5124, 2017.
- Michio Ono, Stefan Kubik, and Chad D Abernathey. *Atlas of the cerebral sulci*. Thieme, 1990.
- Gheorghe Postelnicu, Lilla Zollei, and Bruce Fischl. Combined volumetric and surface registration. *IEEE transactions on medical imaging*, 28(4):508–522, 2008.
- Armin Raznahan, Phillip Shaw, Francois Lalonde, Mike Stockman, Gregory L Wallace, Dede Greenstein, Liv Clasen, Nitin Gogtay, and Jay N Giedd. How does your cortex grow? *Journal of Neuroscience*, 31(19):7174–7177, 2011.
- Emma C Robinson, Saad Jbabdi, Matthew F Glasser, Jesper Andersson, Gregory C Burgess, Michael P Harms, Stephen M Smith, David C Van Essen, and Mark Jenkinson. Msm: a new flexible framework for multimodal surface matching. *Neuroimage*, 100:414–426, 2014.
- Emma C Robinson, Kara Garcia, Matthew F Glasser, Zhengdao Chen, Timothy S Coalson, Antonios Makropoulos, Jelena Bozek, Robert Wright, Andreas Schuh, Matthew Webster, et al. Multimodal surface matching with higher-order smoothness constraints. *Neuroimage*, 167:453–465, 2018.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- Andreas Schuh, Antonios Makropoulos, Robert Wright, Emma C Robinson, Nora Tusor, Johannes Steinweg, Emer Hughes, Lucilio Cordero Grande, Anthony Price, Jana Hutter,

- et al. A deformable model for the reconstruction of the neonatal cortex. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, pages 800–803. IEEE, 2017.
- Mohamed A Suliman, Logan ZJ Williams, Abdulah Fawaz, and Emma C Robinson. A deep-discrete learning framework for spherical surface registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 119–129. Springer, 2022.
- Mohamed A Suliman, Logan ZJ Williams, Abdulah Fawaz, and Emma C Robinson. Unsupervised multimodal surface registration with geometric deep learning. *Medical Image Analysis*, page 103821, 2025.
- Paul M Thompson, Craig Schwartz, Robert T Lin, Aelia A Khan, and Arthur W Toga. Three-dimensional statistical analysis of sulcal variability in the human brain. *Journal of Neuroscience*, 16(13):4261–4274, 1996.
- Logan ZJ Williams, Sean P Fitzgibbon, Jelena Bozek, Anderson M Winkler, Ralica Dimitrova, Tanya Poppe, Andreas Schuh, Antonios Makropoulos, John Cupitt, Jonathan O’Muircheartaigh, et al. Structural and functional asymmetry of the neonatal cerebral cortex. *Nature human behaviour*, 7(6):942–955, 2023.
- Jianxiao Wu, Gia H Ngo, Douglas Greve, Jingwei Li, Tong He, Bruce Fischl, Simon B Eickhoff, and BT Thomas Yeo. Accurate nonlinear mapping between mni volumetric and freesurfer surface coordinate systems. *Human brain mapping*, 39(9):3793–3808, 2018.
- BT Thomas Yeo, Mert R Sabuncu, Tom Vercauteren, Nicholas Ayache, Bruce Fischl, and Polina Golland. Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*, 29(3):650–668, 2009.
- BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 2011.
- Fenqiang Zhao, Zhengwang Wu, Fan Wang, Weili Lin, Shunren Xia, Dinggang Shen, Li Wang, and Gang Li. S3reg: superfast spherical surface registration based on deep learning. *IEEE transactions on medical imaging*, 40(8):1964–1976, 2021.
- Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1529–1537, 2015.

Appendix A.

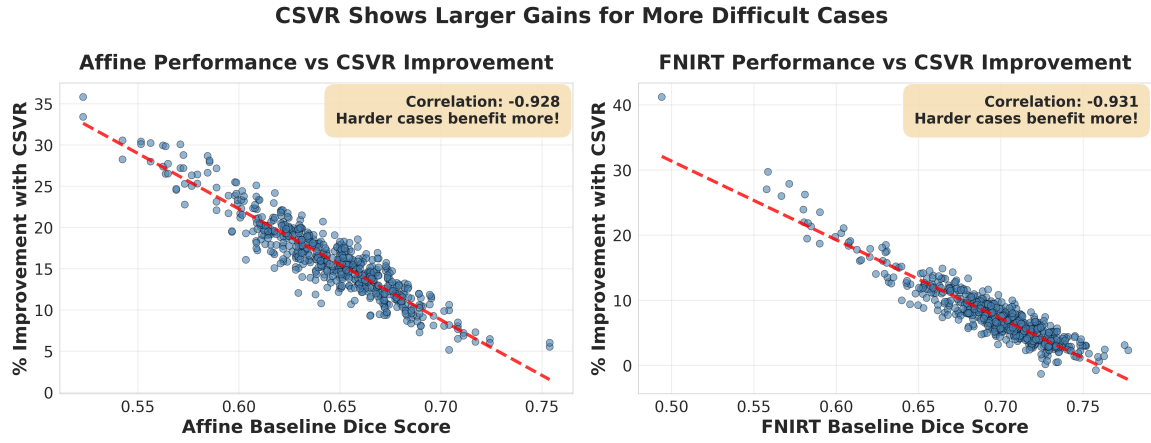


Figure A.1: Performance of other models (Affine and FNIRT) vs % Improvement with CSVr. Harder cases for which FNIRT and affine alignment is poor, benefit more from CSVr registration.