8. Supplementary Material

8.1. Adjoint sensitivity method (ASM) for neural ODE optimization

Here we briefly introduce the Adjoint sensitivity method (ASM) for neural ODE optimization. While the loss function $\mathcal{L}$ in Eqn (4) can be any differentiable function, we will describe ASM by assuming $\mathcal{L}$ to be the mean squared error (MSE) between the resulting flow $\mathrm{z}(t_1)$ and the label z_l which is given by (4). The only reason we express MSE in this form is that it's more convenient for proving ASM convergence. We can therefore formulate the following optimization problem

$$\begin{aligned} \min_{\theta} \quad & \mathcal{L}(\mathrm{z}(t_1)) = \int_{t_0}^{t_1} \delta(t_1 - t) \|\mathrm{z}(t) - \mathrm{z}_l\|_2^2 dt, \\ \text{s.t.} \quad & \frac{d\mathrm{z}}{dt} = f_\theta(\mathrm{z}(t), t), \\ & \mathrm{z}(t_0) = \mathrm{z}_0, \end{aligned} \tag{13}$$

where $\delta(\cdot)$ is the Dirac delta function which ensures that only the gradients of the loss function with respect to z at $t = t_1$ gets propagated back. To propagate the gradient from the loss function to the parameters θ, first we numerically solve the differential equation $d\mathrm{z}/dt = f_\theta(\mathrm{z}(t), t)$ for its trajectory forwards in time from t_0 to t_1 with the initial condition $\mathrm{z}(t_0) = \mathrm{z}_0$. Then we can define the adjoint equation given by:

$$\frac{d\lambda^T}{dt} = -\lambda^T \frac{\partial f}{\partial \mathrm{z}}\bigg|_{\mathrm{z}=\mathrm{z}(t)} + \frac{\partial \mathcal{L}}{\partial \mathrm{z}}\bigg|_{\mathrm{z}=\mathrm{z}(t)}, \tag{14}$$

where λ^T is a continuous-time Lagrange multiplier, also known as the adjoint variable. In the case of MSE, we have $\partial\mathcal{L}/\partial z = 2 \cdot \delta(t_1 - t)(\mathrm{z}(t) - z_l)$. We then numerically solve this equation backwards in time with the initial condition $\lambda^T(t_1) = 0$ to obtain the trajectory of λ^T from $t = t_1$ to $t = t_0$. Lastly the gradient of the loss function with respect to the parameters, also known as the sensitivity is given by

$$\frac{d\mathcal{L}}{d\theta} = -\int_{t_0}^{t_1} \lambda^T \frac{\partial f}{\partial \theta}\bigg|_{\theta=\theta(t)} dt. \tag{15}$$

After obtaining this gradient, we can then perform optimization with methods such as gradient descent. Note that the Jacobians $\partial f/\partial \mathrm{x}$ and $\partial f/\partial \theta$ can be computed efficiently using automatic differentiation during the forward pass. In summary, ASM solves for gradients through the following steps:

- Numerically solve $d\mathrm{z}/dt = f_\theta(\mathrm{z}, t)$ forward in time from t_0 to t_1.
- Numerically solve the adjoint equation (14) backward in time from t_1 to t_0 using the initial condition $\lambda^T(t_1) = 0$.
- Numerically evaluate the integral in Eqn. (15) to obtain the desired gradient.

8.2. Illustrative Examples in 2D pair images: results of LDDMM

To explore the different properties of solution transformations between LDDMM and ours, we conduct the same experiment on 2D pair examples as in Figure 1 using LDDMM shooting method. We use the Mermaid registration toolkit*, with a learning rate of 0.01 and 500 epoches of optimization. The results of LDDMM are shown in Figure 6. LDDMM generates a smoother transformation field, while our method produces a more accurate match without violating diffeomorphism.

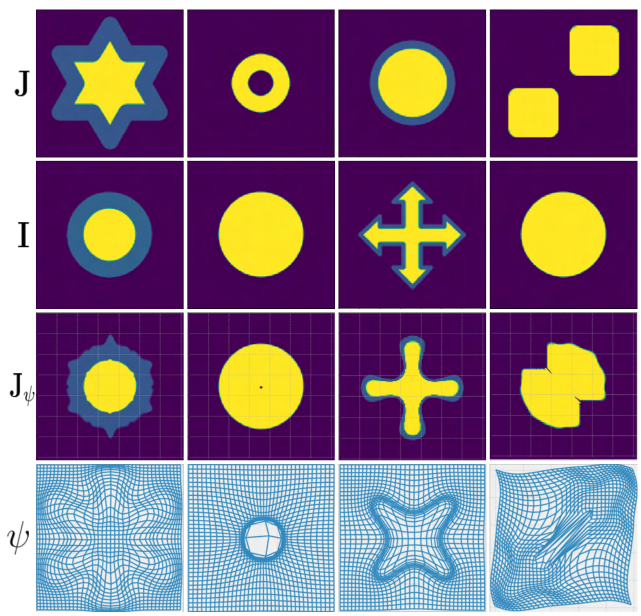

Figure 6. The rows show (J) the moving images, (I) fixed images, (J_ψ) warped moving images, and visualization of ψ respectively.

8.3. Anatomical structures

The details of 28 anatomical structures on which dice scores are calculated are provided in Figure 7. For both the OASIS and CANDI datasets, our method demonstrates a consistent and significant improvement in both mean dice scores over the all brain structures and on each anatomical category as shown in Figure 7.

8.4. Qualitative comparison

We present the qualitative results of SYMNet and ours in Figure 5. In this supplementary material, we provide qualitative comparisons of full benchmarks in Figure 8.

*https://github.com/uncbiag/mermaid

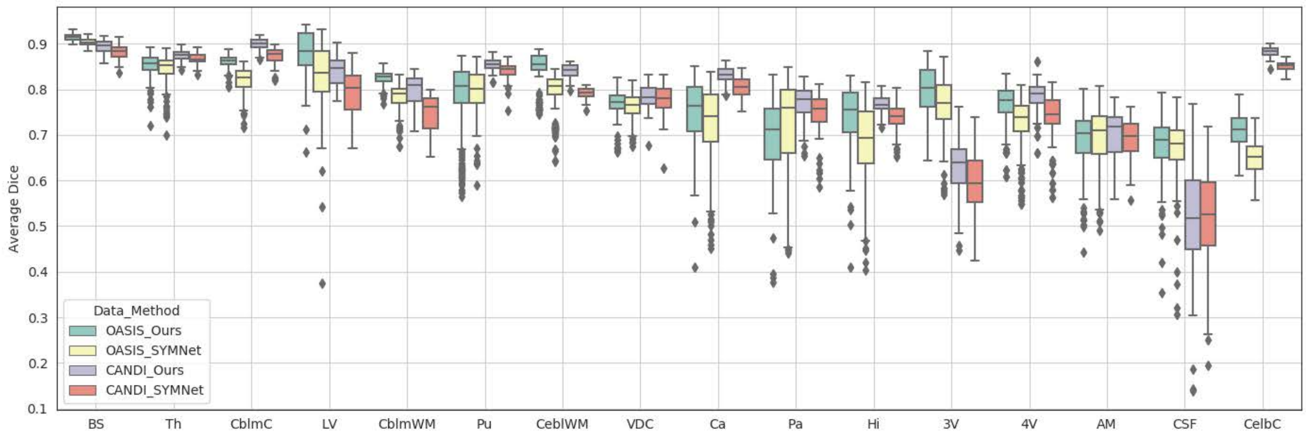

Figure 7. Boxplots indicating Dice for 28 anatomical structures on OASIS and CANDI datasets for SYMNet and our method. The abbreviations here represent brain stem (BS), thalamus (Th), cerebellum cortex (CblmC), lateral ventricle (LV), cerebellum white matter (CblmWM), putamen (Pu), cerebral white matter (CeblWM), Ventral DC (VDC), caudate (Ca), pallidum (Pa), hippocampus (Hi), 3rd ventricle (3V), 4rd ventricle (4V), amygdala (AM), CSF (CSD), cerebral cortex (CelbC).

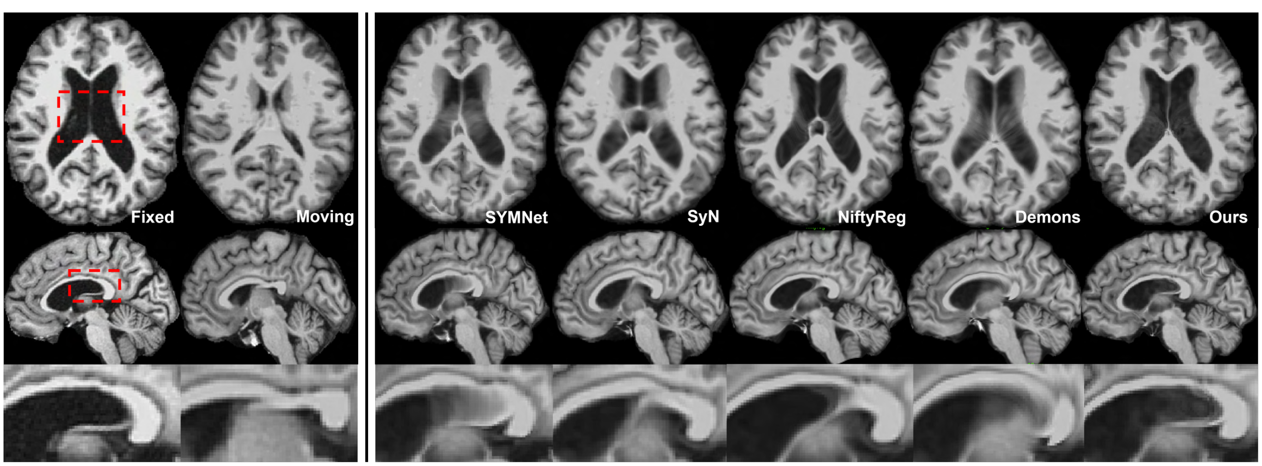

Figure 8. Images showing an example of a registration image pair. Fixed image is OASIS ID001 and Moving image is OASIS ID002. The 3rd column to the 7th column are results of SYMNet, SyN, NiftyReg, Log-Demons and ours respectively.