Supplemental Materials

1   Derivation of evidence lower bound (ELBO)

In the following, we show how to derive the evide lower bound (ELBO) from Eq. 3 of the main text, starting from the KL-divergence between approximate posterior $q(\mathbf{z}|\mathbf{m}, \mathbf{f})$ and the true posterior $p(\mathbf{z}|\mathbf{m}, \mathbf{f})$:

$$
\begin{aligned}
\mathbf{KL}[q(\mathbf{z}|\mathbf{m}, \mathbf{f})||p(\mathbf{z}|\mathbf{m}, \mathbf{f} &= \mathbf{E}_q[\log q(\mathbf{z}|\mathbf{m}, \mathbf{f}) - \log p(\mathbf{z}|\mathbf{m}, \mathbf{f})] \\
=&\mathbf{E}_q\left[\log \prod_{l=0}^{L-1} q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) - \log \prod_{l=0}^{L-1} p(\mathbf{z}_l|\mathbf{m}, \mathbf{f})\right] \\
=&\mathbf{E}_q\left[\log \prod_{l=0}^{L-1} q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) - (\log \prod_{l=0}^{L-1} p(\mathbf{z}_l, \mathbf{m}, \mathbf{f}) - \log p(\mathbf{m}, \mathbf{f}))\right] \\
=&\mathbf{E}_q\left[\log \prod_{l=0}^{L-1} q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) - \log \prod_{l=0}^{L-1} p(\mathbf{z}_l, \mathbf{m}, \mathbf{f})\right] + \log p(\mathbf{m}, \mathbf{f}) \\
=&\mathbf{E}_q\left[\log \prod_{l=0}^{L-1} q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) - (\log p(\mathbf{f}|\mathbf{z}, \mathbf{m}) + \log \prod_{l=0}^{L-1} p(\mathbf{z}_l) + \log p(\mathbf{m}))\right] + \log p(\mathbf{m}, \mathbf{f}) \\
=&\mathbf{E}_q\left[\sum_{l=0}^{L-1}[\log q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) - \log p(\mathbf{z}_l)] - \log p(\mathbf{f}|\mathbf{z}, \mathbf{m}) - \log p(\mathbf{m})\right] + \log p(\mathbf{m}, \mathbf{f}) \\
=&\sum_{l=0}^{L-1}[\mathbf{KL}(q(\mathbf{z}_l|\mathbf{z}_{>l}, \mathbf{m}, \mathbf{f})|p(\mathbf{z}_l))] - \mathbf{E}_q[\log p(\mathbf{f}|\mathbf{z}, \mathbf{m})] + \log p(\mathbf{m}, \mathbf{f}) - \log p(\mathbf{m}) \, .
\end{aligned}
\tag{1}
$$

As $\log p(\mathbf{m}, \mathbf{f})$ and $\log p(\mathbf{m})$ are constant with respect to $\mathbf{z}$, they can not be optimized. The optimization task is thus minimizing the Evidence Lower Bound (ELBO), consisting of the KL-divergence between the approximate posterior $q(\mathbf{z}_l|\mathbf{m}, \mathbf{f})$ and the prior $p(\mathbf{z}_l)$, for each level l and minimizing the negative expectation of the log-likelihood $\log p(\mathbf{f}_l|\mathbf{m}, \mathbf{z}_l)$:

$$
\min \sum_{l=0}^{L-1} \text{KL}[q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f})|p(\mathbf{z}_l)] - \mathbf{E}_q[\log p(\mathbf{f}|\mathbf{z}, \mathbf{m})] \, .
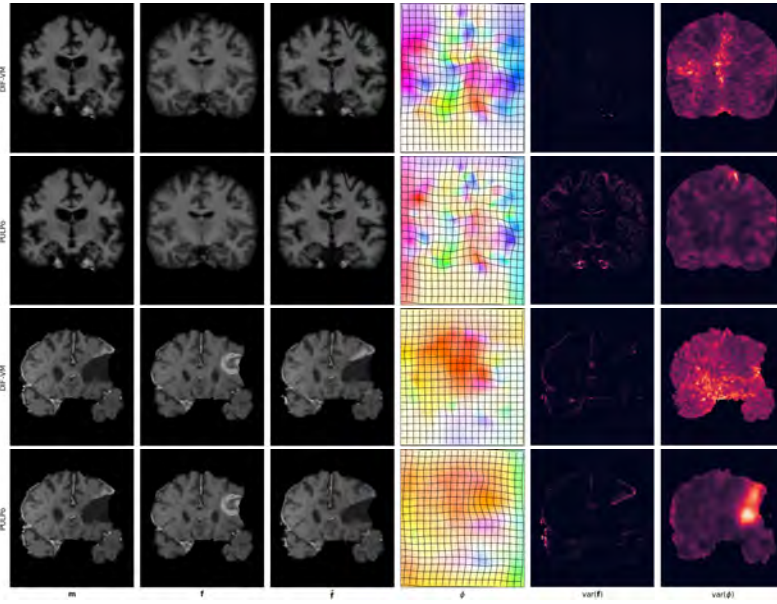\tag{2}
$$

2   Supplemental Figures

Fig. A. Registration Results: The moving image $\mathbf{m}$, fixed image $\mathbf{f}$, prediction $\hat{\mathbf{f}}$, predicted DF ϕ, voxel variance $var(\mathbf{f})$ and DF variance $var(\phi)$. **Above:** Coronal slice 112 of subject 10 ($\mathbf{m}$) and subject 86 ($\mathbf{f}$) on OASIS-1. **Below:** Coronal slice 134 of T1CE-scan 7 on BraTS-Reg.
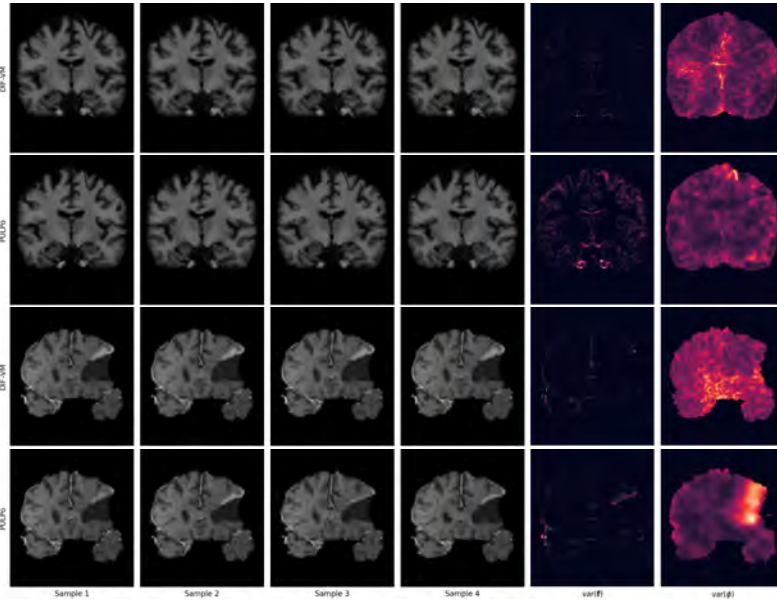

Fig. B. Sample Diversity: Showing 4 samples, the image variance, and the DF variance. **Above:** Coronal slice 112 of subject 10 ($\mathbf{m}$) and subject 86 ($\mathbf{f}$) on OASIS-1. **Below:** Coronal slice 134 of T1CE-scan 7 on BraTS-Reg.