MultiMorph: On-demand Atlas Construction

Supplementary Material

6. Ablation Studies

We conduct several ablations to quantify the effect of individual components of the proposed model.

6.1. Effect of Synthetic Data

We first evaluate the effect of training with and without synthetic data. We presents results on the generalization experiment of Section 4.2.1. We evaluate on the held-out IXI dataset, quantifying the results on T1-w, T2-w, and PD-w image modalities. Table 6 presents the results. In all cases, the inclusion of synthetic data improves the segmentation transfer performance with negligible increase in centrality and number of folds.

6.2. Model ablations

We quantify the effect of several key model components on the OASIS-1 dataset, as described in Section 4.3. Here, we assess the effect on subgroup atlas construction.

Subgroup Atlas Construction. We hypothesize that constructing atlases for homogeneous groups benefits more from within-group feature interactions than heterogeneous groups, by capturing set-specific information. To test this hypothesis, we split the OASIS-1 test set into random subgroups of $[5, 10, 20, 30, 40]$ images and quantify performance. Figure 7 presents the results on the segmentation transfer task. Table 5 presents average results across all subgroups. The effect of the GroupBlock mechanism is immediately apparent, leading to a large increase in Dice score while maintaining well-behaved deformation fields. The improvement enabled by the Group Block mechanism is especially evident in homogeneous groups. For narrow atlas construction tasks, feature sharing within an image group is helpful to produce meaningful, group-specific atlases.

Table 5. Model subgroup ablations. We aggregate performance on atlases created from random subgroups of [5,10,20,30,40] images from the OASIS-1 test set. The GB effectively shares group features, improving subgroup atlas construction.

Ablation	Dice (↑)	Folds (↓)	Centrality $\times 10^{-3}$ (↓)
GB (mean)+Dice	$\mathbf{0.911 \pm 0.002}$	7.1 ± 1.4	18.7 ± 0.5
GB (mean)	0.879 ± 0.005	0.7 ± 0.4	13.8 ± 1.4
GB (max)	0.878 ± 0.005	0.8 ± 0.4	14.3 ± 1.2
GB (var)	0.878 ± 0.006	0.7 ± 0.3	14.0 ± 1.3
no GB	0.862 ± 0.006	$\mathbf{0.0 \pm 0.0}$	$\mathbf{12.4 \pm 2.5}$

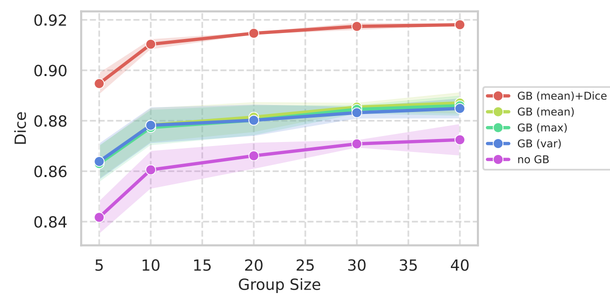

Figure 7. Subgroup atlas construction results across ablation studies on the GroupBlock mechanism. Shaded regions denote the 95% confidence interval. Including the GB mechanism led to significant improvements in segmentation transfer compared to without. Further, training with the Dice loss led to a consistent improvement of up to 2 Dice points.

7. Sensitivity Analysis

We quantify the sensitivity of our model performance to hyperparameters. Using the OASIS-1 validation set, we measure the effect of changing the regularization hyperparameter λ and the Dice loss hyperparameter γ in the produced atlas. Specifically, we measure the effect on Dice transfer, number of folds, and Centrality.

Figure 8 shows results while varying λ and setting $\gamma = 0$. We observe well behaved deformation fields with strong structural alignment for $\lambda \in [0.5, \ldots, 2]$, indicating our model is robust to the choice of this hyperparameter. We set $\lambda = 1$ for all experiments as it achieves a good trade-off between structural alignment and smooth deformation fields.

Figure 9 shows performance while varying γ and setting $\lambda = 1$. The model shows some sensitivity to the Dice loss weight, though maintains strong performance for $\gamma \in [0.1, \ldots, 0.7]$. We select $\gamma = 0.5$ and $\lambda = 1$ for all experiments in the paper. This set of hyperparameters achieved a reasonable tradeoff between structural matching while maintaining regular and smooth deformation fields.

8. Additional Qualitative Results

We present additional qualitiative results of our produced atlases. Figure 10 presents example images and warps to the whole-population IXI atlases. Examples are presented for the T1-w, T2-w and PD-w modalities. Despite differences in contrast and image quality, our single model is able to successfully map individual images to the constructed atlases.

Table 6. IXI held out dataset atlas construction results, comparing our method trained with and without synthetic data.

Modality	**Method**	**Dice Transfer** (↑)	**Folds** (↓)	**Norm Disp.** (↓)	**Centrality** $\times 10^{-3}$ (↓)
T1-w	Ours (w/ Synth)	$\mathbf{0.911 \pm 0.007}$	1.1 ± 1.634	1.659 ± 0.204	13.5 ± 40.914
	Ours (no Synth)	0.894 ± 0.011	0.5 ± 1.057	1.552 ± 0.171	10.0 ± 29.452
T2-w	Ours (w/ Synth)	0.904 ± 0.008	1.7 ± 2.346	1.74 ± 0.209	13.7 ± 40.101
	Ours (no Synth)	0.888 ± 0.013	0.7 ± 1.295	1.611 ± 0.181	8.9 ± 24.201
PD-w	Ours (w/ Synth)	0.897 ± 0.011	0.6 ± 1.299	1.599 ± 0.205	8.9 ± 27.473
	Ours (no Synth)	0.882 ± 0.015	0.3 ± 1.176	1.491 ± 0.172	6.5 ± 19.39

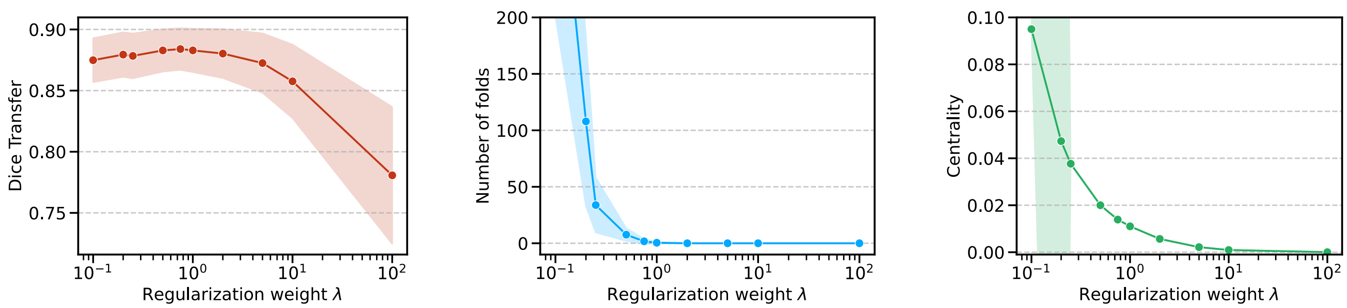

Figure 8. Hyperparameter sweep over regularization weight λ with Dice loss weight $\gamma = 0$ on the OASIS-1 validation set. Shaded regions represent one standard deviation from the mean. Plots show the effect on Dice segmentation transfer, number of folded voxels, and Centrality. Our model shows consistent performance for $\lambda \in [0.5, \ldots, 2]$, indicating robustness. We select $\lambda = 1$ as it achieves a reasonable tradeoff between segmentation alignment and field regularity.

Figure 11 presents examples of synthetic images used in training. The variety of imaging contrasts sampled aids our model's ability to generalize to unseen modalities.

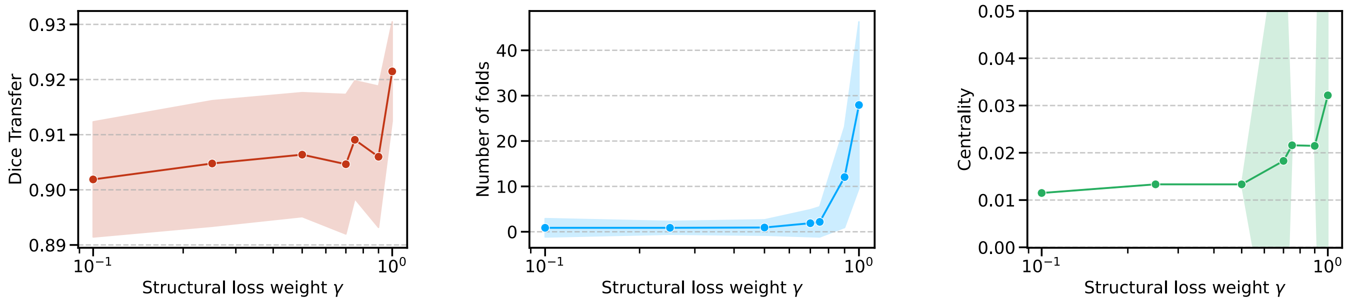

Figure 9. Hyperparameter sweep over Dice loss weight γ with regularization weight $\lambda = 1$ on the OASIS-1 validation set. Shaded regions represent one standard deviation from the mean. Plots show the effect on Dice segmentation transfer, number of folded voxels, and Centrality. Our model shows some sensitivity but achieves consistent performance for $\gamma \in [0.1, \ldots, 0.7]$. We select $\gamma = 0.5$ as it achieves strong segmentation performance while maintaining well-behaved deformation fields.

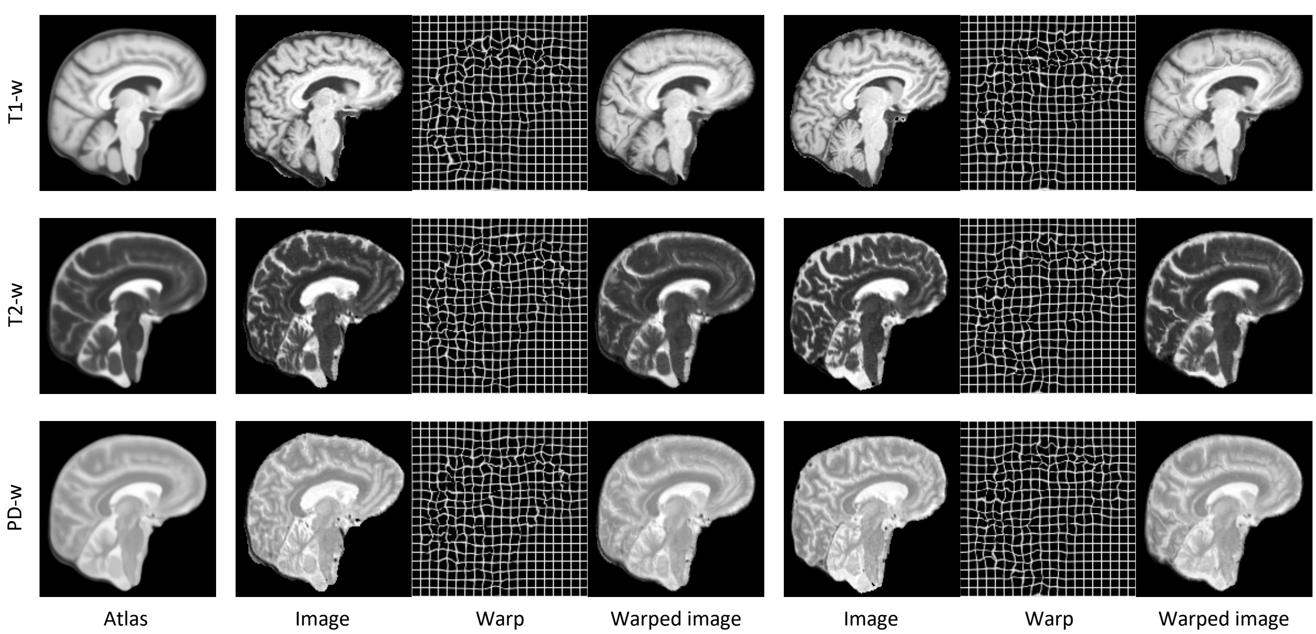

Figure 10. Example images and warps produced by our model on the IXI dataset.

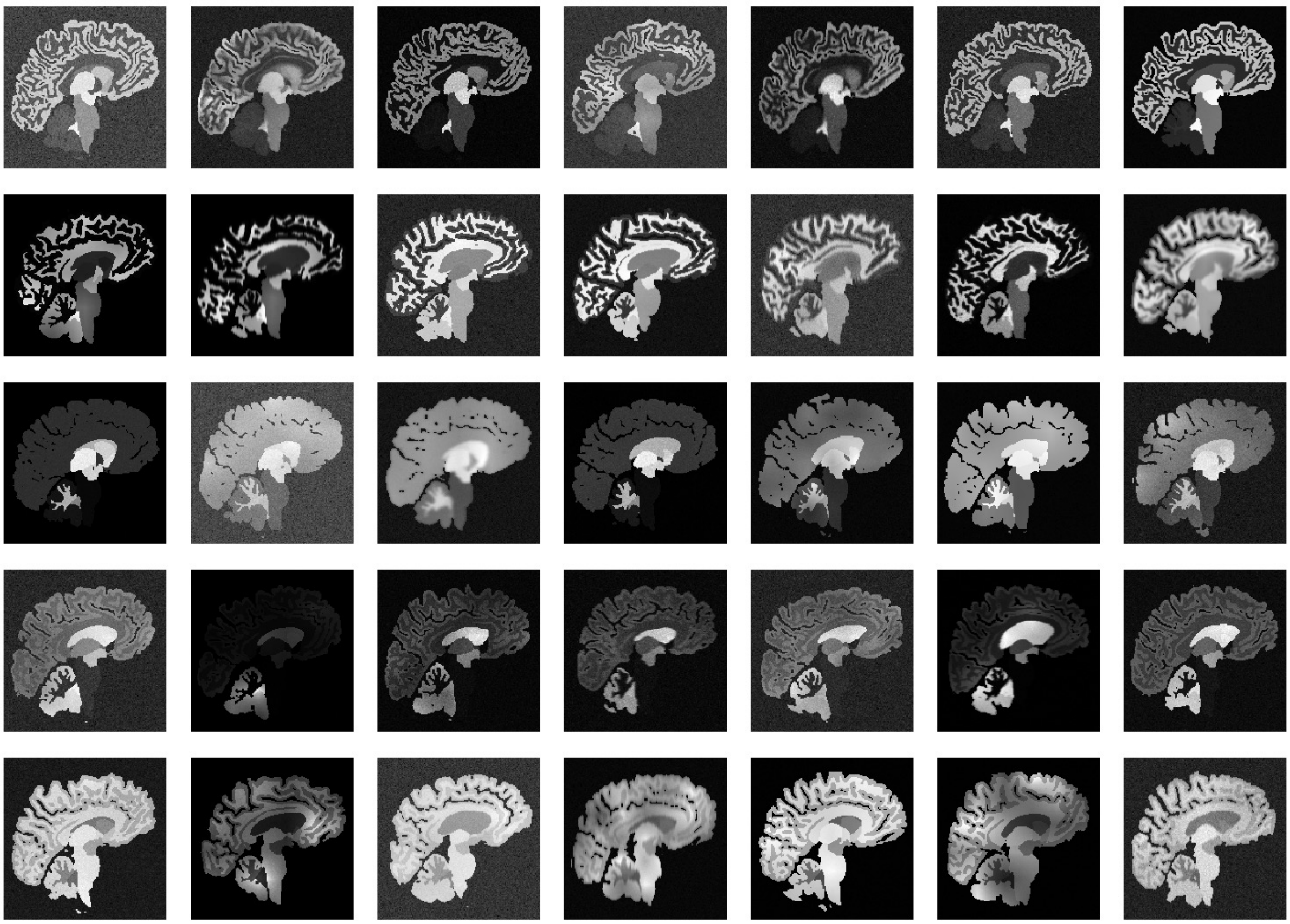

Figure 11. Example synthetic images used in training. Each row represents one group sampled from the same distribution of image contrast, with augmentations performed.