HyperMorph: Amortized Hyperparameter Learning for Image Registration

Andrew Hoopes[1], Malte Hoffmann[1,2], Bruce Fischl[1−3],
John Guttag[3], and Adrian V. Dalca[1−3]

[1] Martinos Center for Biomedical Imaging, MGH
[2] Department of Radiology, Harvard Medical School
[3] Computer Science and Artificial Intelligence Lab, MIT

Abstract. We present HyperMorph, a learning-based strategy for deformable image registration that removes the need to tune important registration hyperparameters during training. Classical registration methods solve an optimization problem to find a set of spatial correspondences between two images, while learning-based methods leverage a training dataset to learn a function that generates these correspondences. The quality of the results for both types of techniques depends greatly on the choice of hyperparameters. Unfortunately, hyperparameter tuning is time-consuming and typically involves training many separate models with various hyperparameter values, potentially leading to suboptimal results. To address this inefficiency, we introduce amortized hyperparameter learning for image registration, a novel strategy to *learn* the effects of hyperparameters on deformation fields. The proposed framework learns a hypernetwork that takes in an input hyperparameter and modulates a registration network to produce the optimal deformation field for that hyperparameter value. In effect, this strategy trains a single, rich model that enables rapid, fine-grained discovery of hyperparameter values from a continuous interval at test-time. We demonstrate that this approach can be used to optimize multiple hyperparameters considerably faster than existing search strategies, leading to a reduced computational and human burden as well as increased flexibility. We also show several important benefits, including increased robustness to initialization and the ability to rapidly identify optimal hyperparameter values specific to a registration task, dataset, or even a single anatomical region, all without retraining the HyperMorph model. Our code is publicly available at `http://voxelmorph.mit.edu`.

Keywords: Deformable Image Registration · Hyperparameter Tuning · Deep Learning · Amortized Learning.

1 Introduction

Deformable image registration aims to find a set of dense correspondences that accurately align two images. Classical optimization-based techniques for image registration have been thoroughly studied, yielding mature mathematical frameworks and widely used software tools [2,4,8,48,52]. Learning-based registration methods employ image datasets to learn a function that rapidly computes the deformation field between image pairs [7,46,50,54,56,57]. These methods involve choosing registration hyperparameters that dramatically affect the quality of the estimated deformation field. Optimal hyperparameter values can differ substantially across image modality and anatomy, and even small changes can have a large impact on accuracy. Choosing appropriate hyperparameter values is therefore a crucial step in developing, evaluating, and deploying registration methods.

Tuning these hyperparameters most often involves grid or random search techniques to evaluate separate models for discrete hyperparameter values (Figure 1). In practice, researchers typically perform a sequential process of optimizing and validating models with a small subset of hyperparameter values, adapting this subset, and repeating. Optimal hyperparameter values are selected based on model performance, generally determined by human evaluation or additional validation data such

Traditional Strategy   **HyperMorph Strategy**
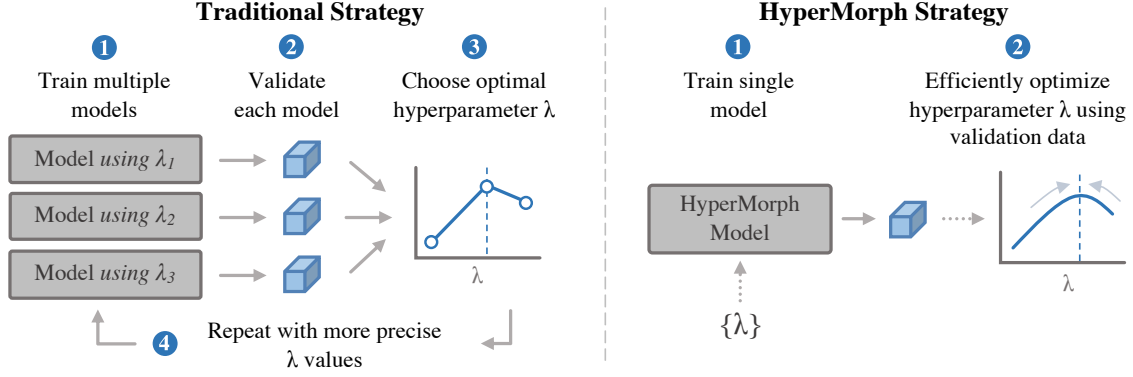
Fig. 1. Hyperparameter optimization strategies. Traditional approaches (left) repeatedly train a registration model, each time with a different hyperparameter value. The proposed HyperMorph approach (right) optimizes a single, richer model once, which approximates a landscape of traditional models.

as anatomical annotations. This approach requires considerable computational and human effort, which may lead to suboptimal parameter choices, misleading negative results, and impeded progress, especially when researchers might resort to using values from the literature that are not adequate for their specific dataset or registration task.

In this work, we introduce a substantially different approach, HyperMorph, to tackle registration hyperparameters: amortized hyperparameter learning for image registration. Our contributions are:

Method. We propose an end-to-end strategy to *learn* the effects of registration hyperparameters on deformation fields with a single, rich model, replacing the traditional hyperparameter tuning process (Figure 1). A HyperMorph model is a hypernetwork that approximates a landscape of registration networks for a range of hyperparameter values, by learning a continuous function of the hyperparameters. Users only need to learn a single HyperMorph model that enables rapid test-time image registration for any hyperparameter value. This eliminates the need to train a multitude of separate models each for a fixed hyperparameter, since HyperMorph accurately estimates their outputs at a fraction of the computational and human effort. In addition, HyperMorph enables rapid, accurate hyperparameter tuning for registration tasks involving many hyperparameters, in which computational complexity renders grid-search techniques ineffective.

Properties. By exploiting implicit weight-sharing, a single HyperMorph model is efficient to train compared to training the many individual registration models it is able to encompass. We also show that HyperMorph is more robust to initialization than standard registration models, indicating that it better avoids local minima while reducing the need to retrain models with different initializations.

Utility. HyperMorph enables rapid discovery of optimal hyperparameter values at *test-time*, either through visual assessment or automatic optimization in the continuous hyperparameter space. We demonstrate the substantial utility of this approach by using a *single* HyperMorph model to identify the optimum hyperparameter values for different datasets, different anatomical regions, or different registration tasks. HyperMorph also offers more precise tuning compared to grid or sequential search.

2   Related Work

Image Registration. Classical approaches independently estimate a deformation field by optimizing an energy function for each image pair. These include elastic models [6], b-spline based deformations [48], discrete optimization methods [16,22], Demons [52], SPM [3], LDDMM [8,13,29,44,58],

DARTEL [2], and symmetric normalization (SyN) [4]. Recent learning-based approaches make use of convolutional neural networks (CNNs) to learn a function that rapidly computes the deformation field for an image pair. Supervised models learn to reproduce deformation fields estimated or simulated by other methods [35,46,50,57], whereas unsupervised strategies train networks that optimize a loss function similar to classical cost functions and do not require the ground-truth registrations needed by supervised methods [7,15,26,34,54].

Generally, all these methods rely on at least one hyperparameter that balances the optimization of an image-matching term with that of a regularization or smoothness term. Additional hyperparameters are often used in the loss terms, such as the neighborhood size of local normalized cross-correlation [5] or the number of bins in mutual information [53]. Choosing optimal hyperparameter values for classical registration algorithms is a tedious process since pair-wise registration typically requires tens of minutes or more to compute. While learning-based methods enable much faster test-time registration, individual model *training* is expensive and can require days to converge, causing the hyperparameter search to consume hundreds of GPU-hours [7,26,54].

Hyperparameter Optimization. Hyperparameter optimization algorithms jointly solve a validation objective with respect to model hyperparameters and a training objective with respect to model weights [21]. The simplest approach treats model training as a black-box function, including grid, random, and sequential search [9]. Bayesian optimization is a more sample-efficient strategy, leveraging a probabilistic model of the objective function to search and evaluate hyperparameter performance [10]. Both approaches are often inefficient, since the algorithms involve repeated optimizations for each hyperparameter evaluation. Enhancements to these strategies have improved performance by extrapolating learning curves before full convergence [19,32] and evaluating low-fidelity approximations of the black-box function [30]. Other adaptations use bandit-based approaches to selectively allocate resources to favorable models [28,36]. Gradient-based techniques differentiate through the nested optimization to approximate gradients as a function of the hyperparameters [38,40,45]. These approaches are computationally costly and require evaluation of a metric on a comprehensive, labeled validation set, which may not be available for every registration task.
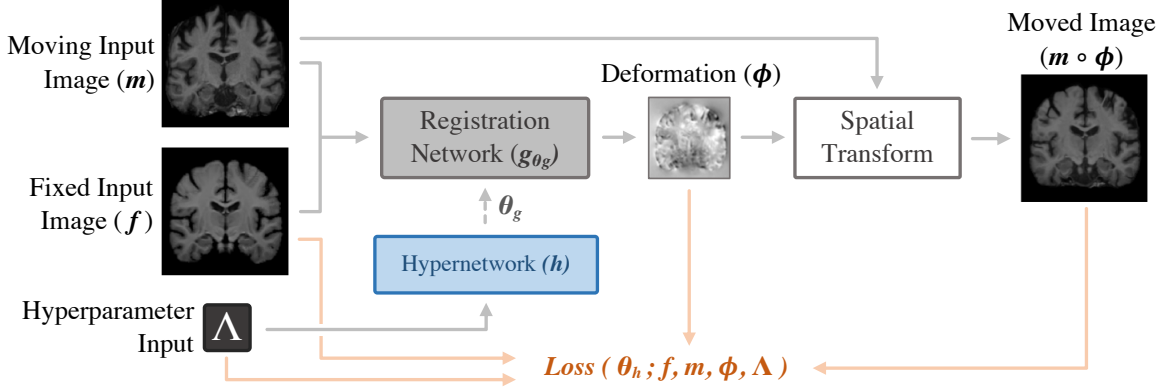
Hypernetworks. Hypernetworks are networks that output weights of a primary network [25,33,49]. Recently, they have gained traction as efficient methods of gradient-based hyperparameter optimization since they enable easy differentiation through the entire model with respect to the hyperparameters of interest. For example, SMASH uses hypernetworks to output the weights of a network conditioned on its architecture [12]. Similar work employs hypernetworks to optimize weight decay in classification networks and demonstrates that sufficiently sized hypernetworks are capable of approximating its global effect [37,39]. HyperMorph extends hypernetworks, combining them with learning-based registration to estimate the effect of hyperparameter values on deformations.

3 Methods

3.1 HyperMorph

Deformable image registration methods find a dense, non-linear correspondence field ϕ between a moving image m and a fixed image f, and can employ a variety of hyperparameters. We follow current unsupervised learning-based registration methods and define a network $g_{\theta_g}(m, f) = \phi$ with parameters θ_g that takes as input the image pair $\{m, f\}$ and outputs the optimal deformation field ϕ.

Our key idea is to model a hypernetwork that learns the effect of *loss* hyperparameters on the desired registration. Given loss hyperparameters Λ of interest, we define the hypernetwork function $h_{\theta_h}(\Lambda) = \theta_g$ with parameters θ_h that takes as input sample values for Λ and outputs the parameters of the registration network θ_g (Figure 2). We learn optimal hypernetwork parameters θ_h

Fig. 2. HyperMorph framework. A hypernetwork (blue) learns to output the parameters of a registration network given registration hyperparameters Λ. HyperMorph is trained end-to-end, exploiting implicit weight-sharing among the full landscape of registration networks within a continuous interval of hyperparameter values.

using stochastic gradient methods, optimizing the loss

$$\mathcal{L}_h(\theta_h; \mathcal{D}) = \mathbb{E}_{\Lambda \sim p(\Lambda)} \Big[\mathcal{L}(\theta_h; \mathcal{D}, \Lambda) \Big], \tag{1}$$

where $\mathcal{D}$ is a dataset of images, $p(\Lambda)$ is a prior probability over the hyperparameters, and $\mathcal{L}(\cdot)$ is a registration loss involving hyperparameters Λ. For example, the distribution $p(\Lambda)$ can be uniform over some predefined range, or it can be adapted based on prior expectations. At every mini-batch, we sample a set of hyperparameter values from this distribution and use these both as input to the network $h_{\theta_h}(\cdot)$ and in the loss function $\mathcal{L}(\cdot)$ for that iteration.

Unsupervised Model Instantiations. Following unsupervised leaning-based registration, we use the loss function:

$$\mathcal{L}_h(\theta_h; \mathcal{D}) = \mathbb{E}_{\Lambda} \Big[\sum_{m,f \in \mathcal{D}} \Big((1-\lambda)\mathcal{L}_{sim}(f, m \circ \phi; \lambda_{sim}) + \lambda \mathcal{L}_{reg}(\phi; \lambda_{reg}) \Big) \Big], \tag{2}$$

where $m \circ \phi$ represents m warped by $\phi = g_{\theta_g}(m, f)$, $\theta_g = h_{\theta_h}(\Lambda)$. The loss term $\mathcal{L}_{sim}$ measures image similarity and might involve hyperparameters λ_{sim}, whereas $\mathcal{L}_{reg}$ quantifies the spatial regularity of the deformation field and might involve hyperparameters λ_{reg}. The regularization hyperparameter λ balances the relative importance of the separate terms, and $\Lambda = \{\lambda, \lambda_{sim}, \lambda_{reg}\}$.

When registering images of the same modality, we use standard similarity metrics for $\mathcal{L}_{sim}$: mean-squared error (MSE) and *local* normalized cross-correlation (NCC). Local NCC includes a hyperparameter defining the neighborhood size. For cross-modality registration, we use normalized mutual information (NMI), which involves a hyperparameter controlling the number of histogram bins [53].

We parameterize the deformation field ϕ with a stationary velocity field (SVF) v and integrate it within the network to obtain a diffeomorphism, which is invertible by design [1,2,15]. We regularize ϕ using $\mathcal{L}_{reg}(\phi) = \frac{1}{2}\|\nabla v\|^2$.

Semi-supervised Model Instantiation. Building on recent learning-based methods that use additional volume information during training [7,26,27], we also apply HyperMorph to the semi-supervised setting by modifying the loss function to incorporate existing training segmentation

maps:

$$\mathcal{L}_h(\theta_h; \mathcal{D}) = \mathbb{E}_\Lambda \sum_{m,f \in \mathcal{D}} \Big[(1-\lambda)(1-\gamma)\mathcal{L}_{sim}(f, m \circ \phi; \lambda_{sim}) \\ + \lambda\mathcal{L}_{reg}(\phi; \lambda_{reg}) + (1-\lambda)\gamma\mathcal{L}_{seg}(s_f, s_m \circ \phi) \Big], \tag{3}$$

where $\mathcal{L}_{seg}$ is a segmentation similarity metric, usually the Dice coefficient [18], weighted by the hyperparameter γ, and s_m and s_f are the segmentation maps of the moving and fixed images, respectively.

3.2   Hyperparameter Tuning

Given a test image pair $\{m, f\}$, a trained HyperMorph model can efficiently yield the deformation field as a function of important hyperparameters. If no external information is available, optimal hyperparameters may be rapidly tuned in an interactive fashion. However, landmarks or segmentation maps are sometimes available for validation subjects, enabling rapid automatic tuning.

Interactive. Sliders can be used to change hyperparameter values in near real-time until the user is visually satisfied with the registration of some image pair $\{m, f\}$. In some cases, the user might choose different settings when studying specific regions of the image. For example, the optimal value of the λ hyperparameter (balancing the regularization and the image-matching term) can vary by anatomical structure in the brain (see Figure 7). This interactive tuning technique is possible because of the HyperMorph ability to efficiently yield the effect of λ values on the deformation ϕ.

Automatic. If segmentation maps $\{s_m, s_f\}$ are available for validation, a single trained HyperMorph model enables hyperparameter optimization using

$$\Lambda^* = \arg\max_\Lambda \mathcal{L}(\Lambda; \theta_h, \mathcal{D}, \mathcal{V}) = \arg\max_\Lambda \sum_{\substack{(m,f) \in \mathcal{D}^2 \\ (s_m, s_f) \in \mathcal{V}^2}} \mathcal{L}_{val}(s_f, s_m \circ \phi), \tag{4}$$

where $\mathcal{V}$ is a set of validation segmentation maps and $\phi = g_{h(\Lambda)}(m, f)$, as before. We implement this optimization by freezing the learned hypernetwork parameters θ_h, treating the input Λ as a parameter to be learned, and using stochastic gradient strategies to rapidly optimize (4).

3.3   Implementation

The hypernetwork we use in the experiments consists of four fully connected layers, each with 64 units and ReLu activation except for the final layer, which uses Tanh activations. The proposed method applies to any registration network architecture, and we treat the hypernetwork and the registration network as a single, large network. The only trainable parameters θ_h are those of the hypernetwork. We implement HyperMorph with the open-source VoxelMorph library [7], using a U-Net-like [47] registration architecture. The U-Net in this network consists of a 4-layer convolutional encoder (with 16, 32, 32, and 32 channels), a 4-layer convolutional decoder (with 32 channels for each layer), and 3 more convolutional layers (of 32, 16, and 16 channels). This results in a total of 313,507 convolutional parameters that are provided as output by the hypernetwork. We use the ADAM optimizer [31] during training.

4   Experiments

We demonstrate that a single HyperMorph model performs on par with and captures the behavior of a rich landscape of individual registration networks trained with separate hyperparameter values,
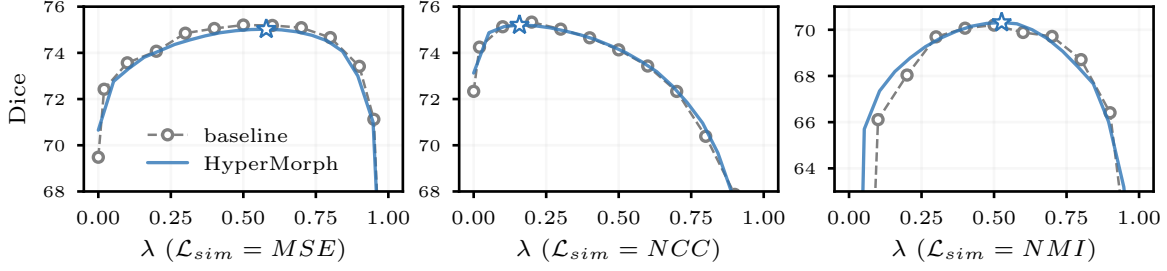
Fig. 3. Mean Dice scores achieved by a single HyperMorph model (blue) and baselines trained for different regularization weights λ (gray) when using each of the MSE, NCC and NMI similarity metrics, respectively. Optima λ^* computed with HyperMorph are indicated by the star markers.

while incurring substantially less computational cost and human effort. We test models with one or two registration hyperparameters. Next, we illustrate considerable improvements in robustness to initialization. Finally, we demonstrate the powerful utility of HyperMorph for rapid hyperparameter optimization at validation — for different subpopulations of data, registration types, and individual anatomical structures.

Datasets. We use two large sets of 3D brain magnetic resonance (MR) images. The first is a multi-site dataset of 30,495 T1-weighted (T1w) scans gathered across 8 public datasets: ABIDE [17], ADHD200 [43], ADNI [55], GSP [14], MCIC [24], PPMI [42], OASIS [41], and UK Biobank [51]. We divide this dataset into train, validation, and test sets of sizes 10,000, 10,000, and 10,495, respectively. The second dataset involves a multi-modal collection of 1,558 T1w, T2-weighted (T2w), multi-flip-angle, and multi-inversion-time images gathered from in-house data and the public ADNI and HCP [11] datasets. We divide this dataset into train, validation, and test sets of sizes 528, 515, and 515, respectively. All MRI scans are conformed to a $256{\times}256{\times}256$ 1-mm isotropic grid space, bias-corrected, and skull-stripped using FreeSurfer [20], and we also produce automated segmentation maps for evaluation. We affinely normalize and uniformly crop all images to $160{\times}192{\times}224$ volumes.

**Evaluation.** For evaluation, we use the volume overlap of anatomical label maps using the Dice metric [18].

Baseline Models. HyperMorph can be applied to any learning-based registration architecture, and we seek to validate its ability to capture the effects of hyperparameters on the inner registration network $g_{\theta_g}(\cdot)$. To enable this insight, we train standard VoxelMorph models with architectures identical to $g_{\theta_g}(\cdot)$ as baselines, each with its fixed set of hyperparameters.

4.1 Experiment 1: HyperMorph Efficiency and Capacity

We aim to evaluate if a single HyperMorph is capable of encapsulating a landscape of baseline models.

Setup. We first assess how the accuracy and computational cost of a single HyperMorph model compare to standard grid hyperparameter search for the regularization weight λ. We separately train HyperMorph as well as VoxelMorph baselines using the similarity metrics MSE (scaled by a constant estimated image noise) and NCC (with window size 9^3) for within-modality registration and NMI (with 32 fixed bins) for cross-modality registration, for which we train 13, 13, and 11 baseline models, respectively. We validate the trained networks on 100 random image pairs for visualization. For hyperparameter optimization after training, we use a subset of 20 pairs.

Additionally, we assess the ability of HyperMorph to learn the effect of multiple hyperparameters simultaneously. We first train a HyperMorph model treating λ and the local NCC window size as
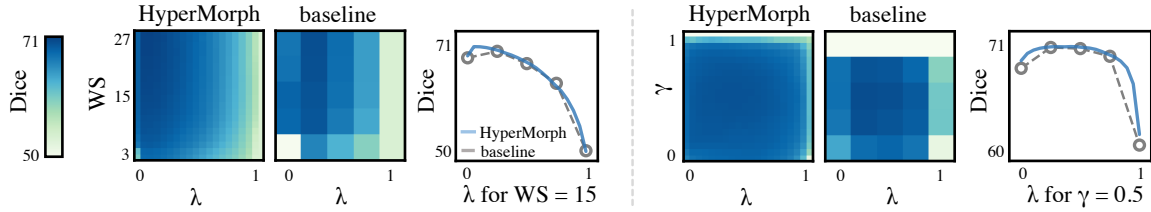
Fig. 4. Two-dimensional hyperparameter search. Left: unsupervised registration with regularization weight λ and local NCC window size WS. Right: semi-supervised registration with hyperparameters λ and segmentation supervision weight γ. For the semi-supervised models, we compute total Dice on both training and held-out labels.

hyperparameters. We also train a semi-supervised HyperMorph model based on a subset of six labels, and hold out six other labels for validation. In this experiment, the hyperparameters of interest are λ and the relative weight γ of the semi-supervised loss (3). Training baselines requires a two-dimensional grid search on 3D models and is computationally prohibitive. Consequently, we conduct these experiments in 2D on a mid-coronal slice, using baselines for 25 combinations of hyperparameter values.

Results. *Computational Cost.* A single HyperMorph model requires substantially less time to convergence than a baseline-model grid search. For single-hyperparameter tests, HyperMorph requires 5.2 ± 0.2 times fewer GPU-hours than a grid search with baseline models (Table 1). For models with two hyperparameters, the difference is even more striking, with HyperMorph requiring 10.5 ± 0.2 times fewer GPU-hours than the baseline models.

Performance. Figures 3 and 4 show that HyperMorph yields optimal hyperparameter values similar to those obtained from a dense grid of baseline models despite the significant computational advantage. An average difference in the optimal hyperparameter value λ^* of only 0.04 ± 0.02 across single-hyperparameter experiments results in a negligible maximum Dice difference of 0.16 ± 0.03 (on a scale of 0 to 100). Similarly, multi-hyperparameter experiments yield a maximum Dice difference of only 0.32 ± 0.02. In practice, fewer baselines might be trained at first for a coarser hyperparameter search, resulting in either suboptimal hyperparameter choice or sequential search leading to substantial manual overhead.

Overall, a *single* HyperMorph model is able to capture the behavior of a range of baseline models individually optimized for different hyperparameters, facilitating optimal hyperparameter choices and accuracy at a substantial reduction in computational cost. We emphasize that the goal of the experiment is not to compare HyperMorph to a particular registration tool, but to demonstrate the effect that this strategy can have on an existing registration network.

Table 1. Comparison between HyperMorph and baseline grid-search techniques for model variability across random initializations (left) and for runtimes (right). We provide runtimes separately for experiments with 1 and 2 hyperparameters (HP).

| | Robustness (init SD) | | Runtime (total GPU-hours) | |
	MSE	NMI	1 HP (3D)	2 HPs (2D)
HyperMorph	**1.97e-1**	**2.46-1**	**146.9 $\pm$ 32.0**	**4.2 $\pm$ 0.6**
Baseline	5.50e-1	5.32e-1	765.3 $\pm$ 249.1	44.0 $\pm$ 4.6
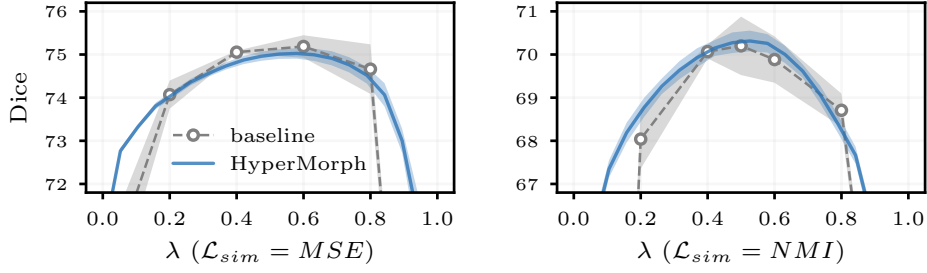
Fig. 5. Variability across several model initializations for HyperMorph and baselines. The shaded areas indicate the standard deviation of registration accuracy, which is substantially more narrow for HyperMorph.

4.2   Experiment 2: Robustness to Initialization

Setup. We evaluate the robustness of each strategy to network initialization. We repeat the previous, single-hyperparameter experiment with MSE and NMI, retraining four HyperMorph models and four *sets* of baselines each trained for five values of hyperparameter λ. For each training run, we re-initialize all kernel weights using Glorot uniform [23] with a different seed. We evaluate each model using 100 image pairs and compare the standard deviation (SD) across initializations of the HyperMorph and baseline networks.

Results. Figure 5 shows that HyperMorph is substantially more robust (lower SD) to initialization compared to the baselines, suggesting that HyperMorph is less likely to converge to local minima. Across the entire range of λ, the average Dice SD for HyperMorph models trained with MSE is 2.79 times lower than for baseline SD, and for NMI-trained models, HyperMorph SD is 2.16 times lower than baseline SD (Table 1). This result further emphasizes the computational efficiency provided by HyperMorph, since in typical hyperparameter searches, models are often trained multiple times for each hyperparameter value to negate potential bias from initialization variability.

4.3   Experiment 3: Hyperparameter-Tuning Utility

Setup. *Interactive Tuning.* We demonstrate the utility of HyperMorph through an interactive tool that enables visual optimization of hyperparameters even if no segmentation data are available. The user can explore the effect of *continuously varying* hyperparameter values using a single trained model and choose an optimal deformation manually at high precision. Interactive tuning can be explored at `http://voxelmorph.mit.edu`.

  Automatic Tuning. When anatomical annotations are available for validation, we demonstrate rapid, automatic optimization of the hyperparameter λ across a variety of applications. In each experiment, we identify the optimal regularization weight λ^* given 20 registration pairs and use 100 registration pairs for evaluation. First, we investigate how λ^* differs across subpopulations and anatomical regions. We train HyperMorph on a subset of image pairs across the entire T1w training set, and at validation we optimize λ separately for each of ABIDE, GSP, PPMI, and UK Biobank. With this same model, we identify λ^* separately for each of 10 anatomical regions. Second, we explore how λ^* differs between cross-sectional and longitudinal registration; for HyperMorph trained on both within-subject and cross-subject pairs from ADNI, we optimize λ separately for validation pairs within and across subjects.

Results. Figures 6 and 7 show that λ^* varies substantially across subpopulations, registration tasks, and anatomical regions. For example, PPMI and ABIDE require a significantly different value of λ^* than GSP and the UK Biobank. Importantly, with a suboptimal choice of hyperparameters, these
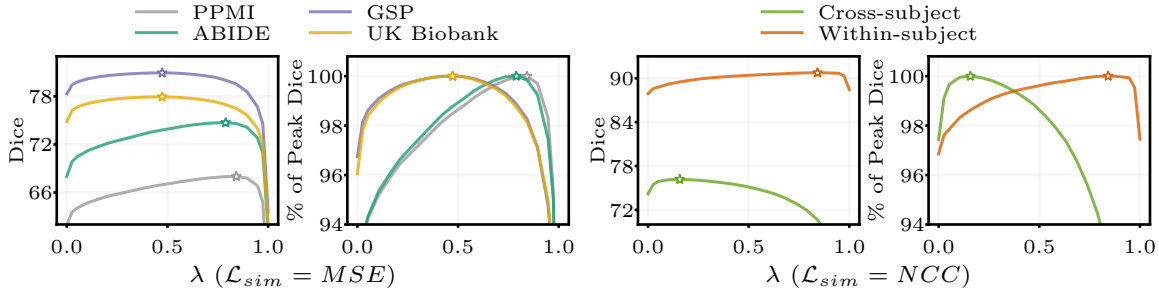
Fig. 6. Registration accuracy across dataset subpopulations (left) and registration tasks (right). The stars indicate the optimal value λ^* as identified by automatic hyperparameter optimization.

datasets would have yielded considerably lower registration quality (Dice scores). The variability in the optimal hyperparameter values is likely caused by differences between the datasets; the average age of the ABIDE population is lower than those of other datasets, while the PPMI scans are of lower quality. Similarly, cross-subject and within-subject registration require different levels of regularization. Finally, Figure 7 illustrates that λ^* varies by anatomical region, suggesting that regularization weights should be chosen by users depending on their tasks downstream from the registration. On average, the automatic hyperparameter optimization takes just 2.8 ± 0.3 minutes using 20 validation pairs.

The vast majority of existing registration pipelines assume a single hyperparameter value to be optimal for an entire dataset, or even across multiple datasets. Our results highlight the importance of HyperMorph as a rapid, easy-to-use tool for finding optimal hyperparameters, interactively or automatically, for different subpopulations, tasks, or even individual anatomical regions, without the need to retrain models.

5   Conclusion

The accuracy of deformable image registration algorithms greatly depends upon the choice of hyperparameters. In this work, we present HyperMorph, a learning-based strategy that removes the need to repeatedly train models to quantify the effects of hyperparameters on model performance. Hyper-Morph employs a hypernetwork which takes the desired hyperparameters as input and predicts the parameters of a registration network tuned to these values. In contrast to existing learning-based
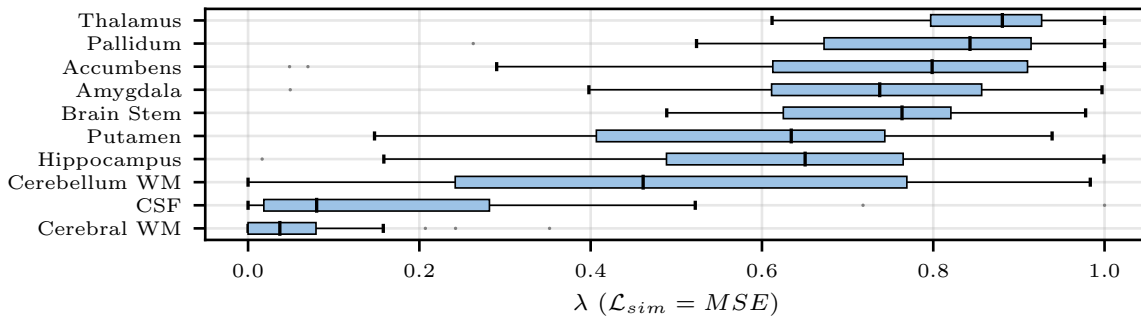
Fig. 7. Optimal regularization weights λ^* across individual anatomical labels. The ranges shown are estimated with HyperMorph for each set of image pairs.

methods, HyperMorph estimates optimal deformation fields for arbitrary image pairs and *any* hyperparameter value from a continuous interval by exploiting sharing of similar weights across the landscape of registration networks. A single HyperMorph model then enables fast hyperparameter tuning at test-time, requiring dramatically less compute and human time compared to existing methods. This is a significant advantage over registration frameworks that are optimized across discrete, predefined hyperparameter values in the hope of finding an optimal configuration.

We demonstrate that a single HyperMorph model facilitates discovery of optimal hyperparameter values for different dataset subpopulations, registration tasks, or even individual anatomical regions. This last result indicates a potential benefit and future direction of estimating a spatially varying field of smoothness hyperparameters for simultaneously optimal registration of all anatomical structures. HyperMorph also provides the flexibility to identify the ideal hyperparameter for an individual image pair. For example, a pair of subjects with very different anatomies would benefit from weak regularization allowing warps of high non-linearity. We are eager to explore hypernetworks for a greater number of hyperparameters. We believe HyperMorph will drastically alleviate the burden of retraining networks with different hyperparameter values, thereby enabling efficient development of finely optimized models for image registration.

Acknowledgements

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: MICCAI: Medical Image Computing and Computer Assisted Interventions. pp. 924–31. Springer (2006)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
3. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. Neuroimage **11**, 805–821 (2000)
4. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. MedIA **12**(1), 26–41 (2008)
5. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
6. Bajcsy, R., Kovacic, S.: Multiresolution elastic matching. Computer Vision, Graphics, and Image Processing **46**, 1–21 (1989)
7. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE TMI **38**(8), 1788–1800 (2019)
8. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV **61**(2), 139–157 (2005)
9. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. JMLR **13**(1), 281–305 (2012)
10. Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. In: NeurIPS. pp. 2546–2554 (2011)
11. Bookheimer, S.Y., Salat, D.H., Terpstra, M., Ances, B.M., Barch, D.M., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Diaz-Santos, M., Elam, J.S., et al.: The lifespan human connectome project in aging: an overview. NeuroImage **185**, 335–348 (2019)

12. Brock, A., Lim, T., Ritchie, J.M., Weston, N.: Smash: one-shot model architecture search through hypernetworks. arXiv preprint arXiv:1708.05344 (2017)
13. Cao, Y., Miller, M.I., Winslow, R.L., Younes, L.: Large deformation diffeomorphic metric mapping of vector fields. IEEE TMI **24**(9), 1216–1230 (2005)
14. Dagley, A., LaPoint, M., Huijbers, W., Hedden, T., McLaren, D.G., Chatwal, J.P., Papp, K.V., Amariglio, R.E., Blacker, D., Rentz, D.M., et al.: Harvard aging brain study: dataset and accessibility. NeuroImage **144**, 255–258 (2017)
15. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med.I.A. **57**, 226–236 (2019)
16. Dalca, A.V., Bobu, A., Rost, N.S., Golland, P.: Patch-based discrete registration of clinical brain images. In: MICCAI: Medical Image Computing and Computer Assisted Interventions PATCHMI. pp. 60–67. Springer (2016)
17. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry **19**(6), 659–667 (2014)
18. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
19. Domhan, T., Springenberg, J.T., Hutter, F.: Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (2015)
20. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
21. Franceschi, L., Frasconi, P., Salzo, S., Grazzi, R., Pontil, M.: Bilevel programming for hyperparameter optimization and meta-learning. arXiv preprint arXiv:1806.04910 (2018)
22. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. MedIA **12**(6), 731–741 (2008)
23. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: AISTATS. pp. 249–256 (2010)
24. Gollub, R.L., Shoemaker, J.M., King, M.D., White, T., Ehrlich, S., Sponheim, S.R., Clark, V.P., Turner, J.A., Mueller, B.A., Magnotta, V., et al.: The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. Neuroinformatics **11**(3), 367–388 (2013)
25. Ha, D., Dai, A., Le, Q.V.: Hypernetworks. arXiv preprint arXiv:1609.09106 (2016)
26. Hoffmann, M., Billot, B., Iglesias, J.E., Fischl, B., Dalca, A.V.: Learning image registration without images (2020)
27. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. MedIA **49**, 1–13 (2018)
28. Jamieson, K., Talwalkar, A.: Non-stochastic best arm identification and hyperparameter optimization. In: AISTATS. pp. 240–248 (2016)
29. Joshi, S.C., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE TIP **9**(8), 1357–1370 (2000)
30. Kandasamy, K., Dasarathy, G., Schneider, J., Póczos, B.: Multi-fidelity bayesian optimisation with continuous approximations. arXiv preprint arXiv:1703.06240 (2017)
31. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
32. Klein, A., Falkner, S., Springenberg, J.T., Hutter, F.: Learning curve prediction with bayesian neural networks (2016)
33. Klocek, S., Maziarka, L., Wolczyk, M., Tabor, J., Nowak, J., Śmieja, M.: Hypernetwork functional image representation. LNCS p. 496–510 (2019)
34. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE TMI **38**(9), 2165–2176 (2019)
35. Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A.: Robust non-rigid registration through agent-based action learning. In: MICCAI: Medical Image Computing and Computer Assisted Interventions. pp. 344–352. Springer (2017)
36. Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A.: Hyperband: A novel bandit-based approach to hyperparameter optimization. JMLR **18**(1), 6765–6816 (2017)

37. Lorraine, J., Duvenaud, D.: Stochastic hyperparameter optimization through hypernetworks. arXiv preprint arXiv:1802.09419 (2018)
38. Luketina, J., Berglund, M., Greff, K., Raiko, T.: Scalable gradient-based tuning of continuous regularization hyperparameters. In: ICML. pp. 2952–2960 (2016)
39. MacKay, M., Vicol, P., Lorraine, J., Duvenaud, D., Grosse, R.: Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions. arXiv preprint arXiv:1903.03088 (2019)
40. Maclaurin, D., Duvenaud, D., Adams, R.: Gradient-based hyperparameter optimization through reversible learning. In: ICML. pp. 2113–2122 (2015)
41. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
42. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al.: The parkinson progression marker initiative (ppmi). Progress in neurobiology **95**(4), 629–635 (2011)
43. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H., et al.: The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in systems neuroscience **6**, 62 (2012)
44. Miller, M.I., Beg, M.F., Ceritoglu, C., Stark, C.: Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. PNAS **102**(27), 9685–9690 (2005)
45. Pedregosa, F.: Hyperparameter optimization with approximate gradient. arXiv preprint arXiv:1602.02355 (2016)
46. Rohé, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: Svf-net: Learning deformable image registration using shape matching. In: MICCAI: Medical Image Computing and Computer Assisted Interventions. pp. 266–274. Springer (2017)
47. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI: Medical Image Computing and Computer Assisted Interventions. pp. 234–241. Springer (2015)
48. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformation: Application to breast mr images. IEEE TMI **18**(8), 712–721 (1999)
49. Schmidhuber, J.: A 'self-referential'weight matrix. In: International Conference on Artificial Neural Networks. pp. 446–450 (1993)
50. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P., Išgum, I., Staring, M.: Nonrigid image registration using multi-scale 3d convolutional neural networks. In: MICCAI: Medical Image Computing and Computer Assisted Interventions. pp. 232–239. Springer (2017)
51. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al.: Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. Plos med **12**(3), e1001779 (2015)
52. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)
53. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. Int J Comput Vis **24**(2), 137–54 (1997)
54. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. MedIA **52**, 128–143 (2019)
55. Weiner, M.W.: Alzheimer's disease neuroimaging initiative (adni) database (2003)
56. Wu, G., Kim, M., Wang, Q., Munsell, B.C., Shen, D.: Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Transactions on Biomedical Engineering **63**(7), 1505–1516 (2015)
57. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration – a deep learning approach. NeuroImage **158**, 378–396 (2017)
58. Zhang, M., Liao, R., Dalca, A.V., Turk, E.A., Luo, J., Grant, P.E., Golland, P.: Frequency diffeomorphisms for efficient image registration. In: IPMI. pp. 559–570. Springer (2017)