LieMorph: Transformer-based Image Registration Using Flows on Lie Groups

Johannes Bostelmann
johannes.bostelmann@uni-luebeck.de
Jan Lellmann
jan.lellmann@uni-luebeck.de

Institute of Mathematics and Image Computing
University of Lübeck
Germany

Abstract

We propose a data-driven approach for deformable image registration that extends the TransMorph architecture with a module based on the flow equation formulated on matrix fields. Choosing a suitable Lie group allows us to precisely control components of the deformation that should be ignored by the regularization and provides a powerful plug-in replacement for the deformation field generator in learning-based methods. Combined with the TransMorph architecture, our method performs favorably on the IXI and OASIS datasets. Our implementation is available at https://github.com/sennhoj321/Liemorph.

1 Introduction

In the scope of this work, we consider methods for solving the *image registration problem:* Given two images modeled as functions $I_1, I_2 : \Omega \to \mathbb{R}$ on an image domain $\Omega \subseteq \mathbb{R}^d$, find a "suitable" *deformation* $\phi : \Omega \to \mathbb{R}^d$ that maps coordinates of points in I_1 to the coordinates of corresponding points in I_2. Image registration is particularly widely used in medical imaging and earth science [17, 24] and is related to other classical computer vision problems such as optical flow estimation and tracking.

Recently, much focus has been on developing data-driven methods based on estimating the deformation using a neural network with the images I_1, I_2 as inputs [15, 28]. A particularly successful implementation is TransMorph [6], which is based on VoxelMorph [3]. TransMorph uses a U-net inspired encoder-decoder structure with Swin Transformer blocks in the encoder [18] and a CNN decoder to generate a dense displacement field from the pair of input images. Most generically, the approach can be viewed as learning the weight parameters θ of a parametrized function f_θ, which takes two discretized images I_i and turns them into a discretized deformation ϕ:

$$\phi = f_\theta(I_1, I_2). \tag{1}$$

However, as already noted by the authors of [3], this makes it hard to control specific properties of the generated deformation, specifically ensuring that it is diffeomorphic. Diffeomorphisms are especially desirable in biomedical applications, where the recovered deformation

should respect the fact that the physical process underlying the motion does not allow tissue to appear, disappear, or self-intersect during motion.

An effective strategy to address this issue was proposed in [3, 6]: Rather than generating the discretized deformation directly, the network is trained to generate coefficients that are fed into a *generator* G that produces the final deformation:

$$\phi = G(f_\theta(I_1, I_2)). \tag{2}$$

By using a generator based on *stationary velocity fields* (SVFs), one can ensure that the resulting deformation is – at least in a spatiotemporally continuous setting – diffeomorphic [1, 2, 12]. Although based on the time-continuous flow equation, the SVF generator can be implemented efficiently using a "scaling and squaring" technique and also allows for an efficient generation of the inverse deformation ϕ^{-1}, for example, to ensure inverse consistency.

However, it remains difficult to precisely control specific components of the deformation, i.e., which image differences should be explained by the deformation and which should be attributed to noise or different image content. Obtaining dense ground truth deformation fields as a reference is notoriously hard [13]. Consequently, a common and successful strategy is to work in a weakly supervised way by 1) including loss terms that assess the plausibility of the deformation based on substitutes such as overlap (Dice scores) between segmented regions, and 2) including *regularization* terms to encourage plausible deformations.

Therefore, finding expressive regularizers and parametrizations is an important task when training such models. When doing so, an important challenge is how to formulate which deformations one wants to categorically allow, i.e., that should be "free" in terms of regularization cost.

Contribution. In this work, we propose *LieMorph*, an architecture and corresponding regularization/training strategy for deformable 3D image registration. Specifically:

- We propose to augment the TransMorph architecture with an SVF-based representation using *matrix-valued velocity fields* inspired by [5], effectively replacing the generator G in (2). Intuitively, we parameterize the deformation $\phi_M(x)$ at a point x as the matrix-vector product of a spatially-varying *matrix field* $M(x)$ and x using homogeneous coordinates. This matrix field is generated from a *matrix-valued velocity field* using an extension of the flow equation, and is restricted to a *matrix group* that is a Lie group, such as the rotational group $\mathrm{SO}(3)$ or special Euclidean group $\mathrm{SE}(3)$.

- The matrix-valued velocity field representation allows to generate non-trivial deformations using *constant* velocity fields. By regularizing the velocity field and suitably choosing the underlying Lie group that defines the allowed matrices $M(x)$, this allows us to precisely control which parts of the deformation should not be penalized – for example, in order not to penalize rotations, one could choose the rotational group.

- We compare our approach with the TransMorph architecture and evaluate it on two tasks: Firstly, in an unsupervised setting on the IXI dataset for atlas-based registration. Secondly, in a weakly supervised setting on the OASIS dataset for inter-patient registration. In both cases, the proposed adaptation with matrix groups improves the performance of the base transformer model (Dice scores in Table 2 and Table 3).

Our implementation is publicly available, see the abstract for the reference.

Further related work. The task of diffeomorphic image registration was formulated as an optimal control problem on a flow equation in [8]. Considering time-variant velocity fields, the authors of [4] formulated a distance on the space of diffeomorphisms in the theory of large deformation diffeomorphic metric mapping (LDDMM). The authors of [2] proposed the scaling-and-squaring approach as a fast solver of the flow equation for the special case of stationary velocity fields. Stationary velocity fields were subsequently used in classical [2, 16, 20, 27] and data-driven approaches for image registration [3, 6, 14]. Matrix groups as deformation modules were used in [21] and combined with stationary velocity fields in [5].

Notable related work in terms of network architecture includes [26], whose authors first proposed the attention module for sequence analysis as an alternative to previous architectures such as recurrent (RNN) and long short-term memory (LSTM) neural networks. Attention modules were adapted for image analysis in the form of shifted window (SWIN) transformers [18]. This module can better adapt to spatially more distant information by substituting the encoder in U-Net architectures as in [3]. Subsequently, SWIN transformers also found use in image registration [6, 22, 23].

Mathematical Preliminaries. A Lie group is a group $(G,\cdot)$ that is also a (topological) manifold, such that the following conditions hold:

- The group multiplication $\cdot : G \times G \to G : (g,h) \mapsto g \cdot h$ is continuous.
- The inversion $()^{-1} : G \to G : g \mapsto g^{-1}$ is continuous.

As a manifold, a tangential space T_pG is assigned to each point $p \in G$. The collection of all these tangential spaces TG is called a tangential bundle. A vector field $X : G \to TG$, $X_p \in T_pG$ on G assigns to each point $p \in G$ a vector X_p in the corresponding tangential space at p. A special case of smooth vector fields is the class of right-invariant (smooth) vector fields, which is of interest for the augmentation of the flow equation. Such vector fields fulfill the criterion of right-invariance:

$$(Dg)_p X_p = X_{p \cdot g} \tag{3}$$

for all $p,g \in G$, where Dg describes the push-forward/differential of $p \mapsto p \cdot g$ [9]. Crucially, these fields can be directly parametrized by their value at the identity and therefore be parametrized with a vector in a possible low-dimensional vector field, coinciding with the dimensionality of the Lie group. The space of right-invariant vector fields will be denoted by $\mathfrak{g}$.

2 Model

We build upon the matrix-valued velocity field approach, which we briefly summarize based on the presentation in the original publication [5]. Our generator G has the following structure: Let $\Omega \subset \mathbb{R}^3$ be a domain, and consider a matrix Lie group G acting on $\mathbb{R}^3$. For a point $x \in \mathbb{R}^3$, we denote its homogeneous coordinate representation by $\bar{x} \in \mathbb{R}^4$, and let $P : \mathbb{R}^4 \to \mathbb{R}^3$ denote the corresponding projection. Rather than specifying the deformation ϕ explicitly, we instead construct a matrix field $M : \Omega \to G$ and use it to define a deformation ϕ_M:

$$\phi_M(x) := PM(x)\bar{x} \in \mathbb{R}^3 \tag{4}$$

As an example, for the special Euclidean group $G := \mathrm{SE}(3)$, the corresponding deformation will be defined by:

$$\phi_M(x) := \begin{pmatrix} I_3 & \mathbf{0} \end{pmatrix} \begin{pmatrix} R(x) & p(x) \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \in \mathbb{R}^3, \tag{5}$$

in which $R(x) \in \mathbb{R}^{3\times 3}$ is a rotation matrix and $p(x) \in \mathbb{R}^3$ a translation vector assigned to the spatial position $x \in \Omega$.

This parameterization allows us to represent not only translations using constant fields M, but also general affine transformations such as rotations, (a) making it potentially easier for the network to generate more complicated deformations, and (b) by only penalizing changes in M instead of ϕ, allowing us to ignore certain deformation components in the regularizer selectively.

Note that the matrix fields M do not have to be constant. Therefore, non-affine deformations can also be constructed in this way, even if the matrix group in question is, for example, $\mathrm{SE}(3)$, the group of roto-translations. In order to generate the matrix field, we follow the SVF approach, i.e., M is obtained by making it depend on an artificially introduced time parameter $t \in [0,1]$, starting with $M(\cdot,0) = id$ at $t = 0$, and integrating the flow equation with a velocity field generated by a neural network to obtain the final matrix field $M(\cdot,1)$. In the classical SVF setting, this allows it to obtain diffeomorphic deformations.

The *flow equation* adapted to matrix groups is [5]

$$\begin{aligned} \frac{\partial M}{\partial t}(x,t) &= v(PM(x,t)\bar{x})_{M(x,t)} \\ M(\cdot,0) &= id, \end{aligned} \tag{6}$$

where the velocity $v : \mathbb{R}^n \to \mathfrak{g}$ functions as the control term of the ODE. A vector field $v(x)$ in $\mathfrak{g}$ evaluated at a matrix $M(x,t) \in G$ is written as $v(x)_{M(x,t)} \in T_{M(x,t)}G$ and lies in the specific tangent space. This velocity vector describes the infinitesimal change of the matrix field M at position x and at time t. After integrating this flow equation for $M(\cdot,1)$, the final deformation is given by $\phi_{M(\cdot,1)}$. The flow equation is solved numerically in our experiments using the scaling-and-squaring technique with $k = 7$ squaring steps. The calculations on the Lie groups can be implemented in a fully differentiable way [25], which makes such a generator accessible to data-driven methods. The approach allows generating any classical stationary velocity field when choosing matrix groups such as $\mathrm{SE}(3)$ and $\mathrm{SIM}(3)$ that include translations. It is therefore a proper generalization of the classical SVF approach.

In this work, we propose to generate the velocity field v by a TransMorph-style network f_θ, which takes a discretized image pair from some image space I as input and produces the velocity $v : \Omega \to \mathfrak{g}$ discretized on a grid, the space of which we denote by V:

$$f_\theta : I \times I \to V. \tag{7}$$

We used the TransMorph architecture [6] as a generator of the velocity fields. The network is designed as a shifted window transformer [18] with sinusoidal positional embedding, consisting of four attention modules as the encoder and a decoder consisting of five convolutional layers. Additional skip connections are added as well. The extension to matrix-valued velocity fields only requires changing the last convolutional layer, whose output channels now depend on the dimensionality of the chosen Lie group.

The dimension of the velocity vectors in $\mathfrak{g}$ mirrors the dimension of the underlying manifold. For our experiments, we chose the Lie groups $\mathrm{SE}(3)$ of roto-translations in the

three-dimensional space, and SIM(3), the Lie group of roto-translations with additional scaling. These groups, considered as manifolds, are six-dimensional for SE(3) and seven-dimensional for SIM(3).

The output is numerically integrated by the scaling-and-squaring approach on the Lie group, which produces the deformation. Finally, this deformation acts as input in the image similarity and segmentation term within the loss. The networks have a size of ~ 0.734 gigabytes in each of our experiments. Training was performed using Adam and a learning rate of 10^{-4}.

Regularization and Loss. During training, the weights are optimized so that the following loss is minimized over image pairs of a given training set:

$$\mathcal{L}(I_1, I_2, ; f_\theta(I_1, I_2)) := D(I_1, I_2 \circ S(f_\theta(I_1, I_2))) + \lambda R_G(f_\theta(I_1, I_2)) + \frac{1}{n} \sum_{i=1}^{n} DS(M_{1,i}, M_{2,i} \circ S(f_\theta(I_1, I_2))). \tag{8}$$

As in [3], the image similarity term D is based on windowed normalized cross-correlation. The third term in (8) is only added in the weakly supervised case where segmentations of I_1 and I_2 are available and measures the Dice score DS between the deformed segmentation masks $M_{j,i}$, $j \in \{1,2\}$, averaged over the n regions.

The crucial new addition is the formulation of the different smoothness term R, which promotes regularity of the deformation and, importantly, now depends on the chosen Lie group, as it is formulated on the underlying velocity field:

$$R_G(v) := \frac{1}{|\Omega|} \int_\Omega \|Dv(x)_{id}\|_F^2 dx \tag{9}$$

where the norm denotes the Frobenius norm of the Jacobian matrix, which defines a linear map $Dv(x) : \mathbb{R}^3 \to TG_{id}$ to the tangential space at the identity. In our experiments, we discretized the differential operator using forward differences. Depending on the chosen Lie group, the kernel of this smoothness term varies. For example, if the Lie group is chosen as SE(3), then constant angular velocities – corresponding to global rotations – are not penalized at all. This differs from the diffusion regularizer used in the original TransMorph method [5],

$$\bar{R}(\phi) := \frac{1}{|\Omega|} \int_\Omega \|D\phi(x)\|_F^2 dx, \tag{10}$$

which penalizes every non-constant displacement field. Note that $\bar{R}$ penalizes deformations while R_G penalizes velocity fields; as the former depends on the latter nonlinearly in the matrix-valued case, there is no simple linear correspondence between the two. This inherent non-linearity also makes balancing the regularization value not a straightforward task. We used empirically $\lambda = 2$ for the Lie group SE(3) and $\lambda = 2.5$ for SIM(3) in all of our registration experiments.

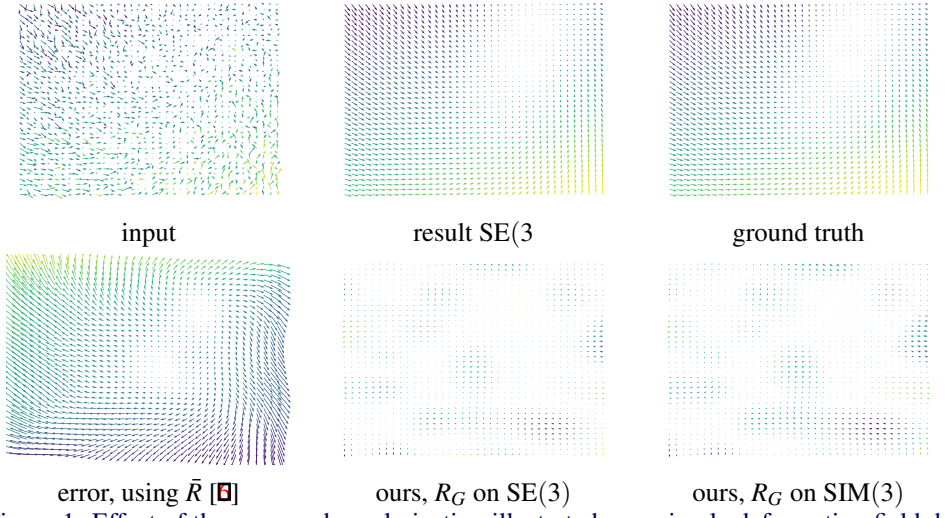

Figure 1: Effect of the proposed regularization illustrated on a simple deformation field denoising task. Shown are 2D slices from the 3D vector fields; the color of the arrows indicates the direction orthogonal to the 2D slice. The ground truth (top right) is dominated by a rotational motion. The proposed velocity representation, here using the Lie group SE(3), allows for accurate recovery of the ground truth from the noisy input (top left) without penalizing roto-translations (top center). In the classical stationary velocity field approach with gradient regularization (bottom left), the error is dominated by a low-frequency component caused by the global rotation. The matrix group representation (bottom center and bottom right) greatly reduces this component and consequently the endpoint error (bottom left).

noise level	SVF	SE(3)	SIM(3)
$\sigma = 0.5$	0.016	0.011	0.010
$\sigma = 2.5$	0.019	0.013	0.012

Table 1: Mean squared error (MSE) to the ground truth for the denoising task (Fig. 1) on the synthetic dataset, averaged over 30 samples each. Regularization weights were $\lambda_1 = 0.1, \lambda_2 = 0.4$, and $\lambda_3 = 0.625$ for the SVF, SE(3), and SIM(3).

Velocity-based regularizers often have the appealing property of inverse-consistency, i.e., the smoothness term of a given deformation coincides with the smoothness term of its inverse. The flow equation ensures that [10]

$$S(v)^{-1} = S(-v), \quad \text{so that} \quad R_G(v) = R_G(-v) \tag{11}$$

ensures inverse consistency. This reduces the influence of order in the registration task, e.g., which image to choose as the reference and template image.

Denoising of synthetic displacement vector fields. To illustrate the impact of the proposed regularization, consider the simple denoising problem in Figure 1, where we used synthetically generated displacement vector fields as ground truth. We chose $7^3 = 343$ equidistantly distributed control points in the image domain $\Omega \subset \mathbb{R}^3$ and applied a rotation around a random axis drawn from a uniform distribution over the unit sphere; the angle of rotation was uniformly distributed over the interval $[-\frac{\pi}{4}, \frac{\pi}{4}]$, producing rotations up to 45 degrees.

method	Dice score (nearest)	Dice score (trilinear)	$\det(J_\phi) \leq 0$
initial	0.406 ± 0.035	0.406 ± 0.035	-
TransMorph [6]	0.744 ± 0.029	0.751 ± 0.030	0.015 ± 0.003
LieMorph SE	0.756 ± 0.025	0.763 ± 0.025	**0**
LieMorph SIM	$\mathbf{0.757 \pm 0.024}$	$\mathbf{0.764 \pm 0.025}$	**0**

Table 2: Unsupervised atlas-based registration on the IXI dataset. Dice scores were averaged over 30 segments and over the image pairs in the test set. The standard deviation was calculated over the segment-wise averaged Dice scores. The LieMorph augmentation improves on the Dice score and drastically reduce the percentage of points where undesirable folding occurs, as indicated by a negative determinant of the Jacobian J_ϕ of the deformation, from an average of 1.5% to zero.

We added a random translation of maximal length $[0, 0.1]$ and an additional non-rigid element generated by perturbing each control point by a random translation with length uniformly distributed over $[0, 0.05]$. The dense synthetic deformation vector field ϕ_{syn} was ultimately computed by interpolating between the perturbed control points using radial basis functions based on second-order polyharmonic splines [12]. We then solved the denoising model

$$J(v;\phi) = \|\phi - S(v)\|^2_{\mathcal{L}^2(\Omega,\mathbb{R}^k)} + \lambda R_G(v), \tag{12}$$

with the regularization term R_G in (9) using implicit neural representations of the velocity fields similar to [5, 14]. We compare the proposed methods using matrix-valued velocity fields with the classical stationary velocity field approach, in which the velocity field acts pointwise as a translation. The results in Figure 1 and Table 1 show that the mean squared error is significantly reduced when using the proposed regularizer.

3 Numerical Results

The numerical experiments were implemented in PyTorch 2.3.0 and performed on a 24-core AMD EPYC 74F3 with three NVIDIA A100 accelerators and CUDA 12.6.

Data Sets. For unsupervised atlas-based registration, we evaluated on the IXI dataset as used in the TransMorph [6] publication. This dataset MRI volumes consist of one atlas image, a training set of size 403, a validation set of size 58, and a test set of size 115; image dimensions are $160 \times 192 \times 224$. More than 40 segmented regions are provided for testing, which were automatically generated using FreeSurfer [11]. Of these segmentations, 30 were chosen for evaluation as in [6], discarding tiny structures.

For weakly supervised training, we used the OASIS-1 dataset [19], which comprises 414 human brain MRI images of patients in different stages of dementia and a healthy control group. We used the skull-stripped, affinely pre-aligned volumes provided in the dataset. The image dimensions are identical to the IXI dataset. In addition, each dataset contains manual segmentations of 35 regions, which were used for computing the Dice scores in the weakly supervised approach (8). The dataset comes split into a training set of size 394 and a validation set of size 19, with no separate test set provided.

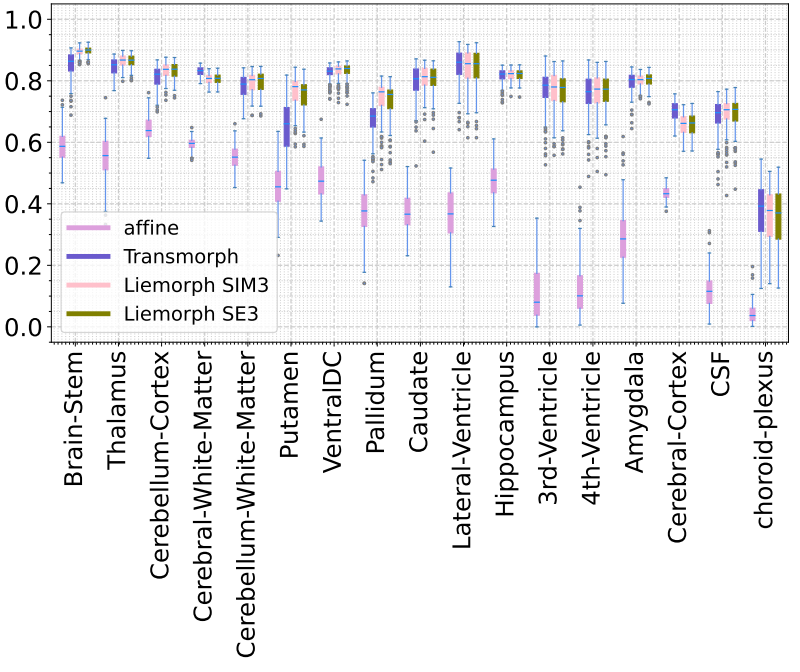

Figure 2: Dice scores for the individual anatomical brain regions on the unsupervised registration task on the IXI dataset. All methods improve on the affine preregistration. While the models using matrix-valued velocity fields improve the overall mean Dice scores compared to TransMorph (Table 2), there are also some structures, such as the cerebral cortex, where the alignment is slightly reduced. The visualization code is based on [6].

method	Dice score (nearest)	Dice score (trilinear)	$\det(J_\phi) \leq 0$
initial	0.5718 ± 0.0517	0.5718 ± 0.0517	-
LieMorph SE	$\mathbf{0.8717 \pm 0.0129}$	$\mathbf{0.8789 \pm 0.0130}$	$1.548 \cdot 10^{-6}$
LieMorph SIM	0.8699 ± 0.0141	0.8772 ± 0.0136	$\mathbf{1.050 \cdot 10^{-6}}$
TransMorph large	0.8623 ± 0.0144	0.8695 ± 0.0146	0.0080
TransMorph	0.8575 ± 0.0145	0.8645 ± 0.0141	0.0084

Table 3: Weakly-supervised inter-patient registration on the OASIS dataset. The Dice score was calculated by taking the mean over the segmented regions of interest and the image pairs in the test set. Augmenting TransMorph with matrix-valued velocity fields also improves Dice scores in this weakly supervised case.

Comparison to TransMorph on IXI. We evaluated the matrix-valued velocity field approach for the task of atlas-based unsupervised registration on the IXI dataset. As the final model, we chose the one with the highest mean Dice score over sets of 30 segmented regions of interest on the validation set. The deformation was evaluated using nearest neighbor interpolation. The test set was not used during training.

The results in Table 2 show that mean Dice scores are improved by combining the matrix-valued velocity fields approach with the TransMorph architecture. As a side note, trilinearly interpolating the binary masks and subsequent rounding noticeably improves the Dice scores compared to simple nearest neighbor interpolation. Therefore, we list both and conclude that being aware of this effect is important when comparing different methods based on published scores.

A closer inspection considering the individual segments is presented in Figure 2. Individually, no model outperforms the other models on all segments under consideration. In contrast to the TransMorph method, our models generate deformations with almost no folding. This is encouraging, as it is one of the main reasons for using classical velocity field-based representations such as SVF and LDDMM, and appears to transfer well to the Lie group setting.

Comparison to TransMorph on OASIS. We evaluated the matrix-valued velocity field for inter-patient semi-supervised registration on the OASIS dataset. The best model during training was chosen based on the mean Dice score of 30 segmented regions of interest on the validation set. In all cases, the deformation was evaluated using nearest neighbor interpolation.

Table 3 shows the results on the validation set when training the method in a weakly supervised way. LieMorph improves on the mean Dice score, outperforming even the larger vanilla TransMorph architecture with a size of ~ 1.78 gigabytes while using the backbone of the standard smaller TransMorph architecture. Additionally, when using our approach, the amount of folding in the generated deformations is again significantly reduced by a factor of over 10^3, indicating that the deformation is numerically "close" to a diffeomorphism.

Experimentally, in addition to achieving higher validation scores, our approach also improves convergence in training compared to the original TransMorph architecture, reducing the number of epochs required to reach a given validation score (Figure 3). While the large TransMorph architecture takes approximately 300 epochs to reach a threshold of 0.86, our LieMorph approach needs less than 70 epochs for both the $\mathrm{SE}(3)$ and $\mathrm{SIM}(3)$ Lie groups.

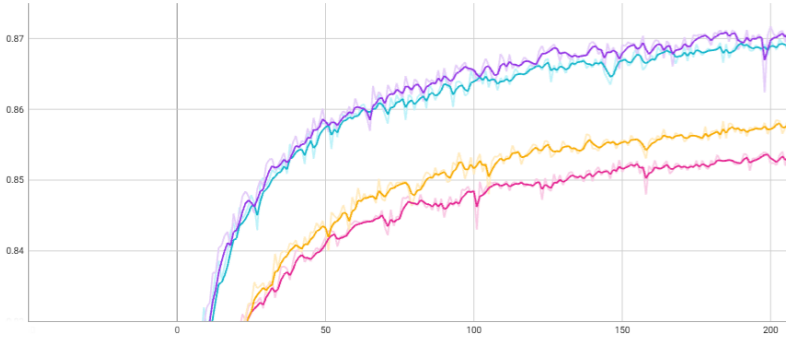

Figure 3: Validation Dice score on the semi-supervised registration task on the OASIS dataset over the training epochs. LieMorph with the matrix groups $\mathrm{SE}(3)$ (cyan) and $\mathrm{SIM}(3)$ (violet) speeds up training compared to plain TransMorph (pink) and TransMorph-large (yellow).

4 Conclusion

Introducing matrix-valued velocity fields on Lie groups into transformer-based data-driven methods appears to be an exciting combination that allows it to fine-tune the desired properties of the deformation field. In the future, it will be particularly interesting to see how these ideas can be extended to the non-stationary (LDDMM) setting and to more expressive groups, such as affine or even projective groups.

References

[1] Sylvain Arguillere, Emmanuel Trélat, Alain Trouvé, and Laurent Younes. Shape deformation analysis from the optimal control viewpoint. *Journal de Mathématiques Pures et Appliquées*, 104(1):139–178, 2015.

[2] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I 9*, pages 924–931. Springer, 2006.

[3] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

[4] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61:139–157, 2005.

[5] Johannes Bostelmann, Ole Gildemeister, and Jan Lellmann. Stationary velocity fields on matrix groups for deformable image registration. *arXiv preprint arXiv:2410.10997*, 2024.

[6] Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis*, 82:102615, 2022.

[7] Zhiye Chen, Lin Li, Jie Sun, and Lin Ma. Mapping the brain in type ii diabetes: voxel-based morphometry using dartel. *European journal of radiology*, 81(8):1870–1876, 2012.

[8] G E Christensen, R D Rabbitt, and M I Miller. 3d brain mapping using a deformable neuroanatomy. *Physics in Medicine & Biology*, 39(3):609, mar 1994. doi: 10.1088/0031-9155/39/3/022. URL https://dx.doi.org/10.1088/0031-9155/39/3/022.

[9] Josi A de Azcárraga and Josi M Izquierdo. *Lie groups, Lie algebras, cohomology and some applications in physics*. Cambridge University Press, 1998.

[10] Paul Dupuis, Ulf Grenander, and Michael I Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, 56(3):587–600, 1998.

[11] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[12] Davide Forti and Gianluigi Rozza. Efficient geometrical parametrisation techniques of interfaces for reduced-order modelling: application to fluid–structure interaction coupling problems. *International Journal of Computational Fluid Dynamics*, 28(3-4):158–169, 2014.

[13] In Young Ha, Matthias Wilms, and Mattias Heinrich. Semantically guided large deformation estimation with deep networks. *Sensors*, 20(5):1392, 2020.

[14] Kun Han, Shanlin Sun, Xiangyi Yan, Chenyu You, Hao Tang, Junayed Naushad, Haoyu Ma, Deying Kong, and Xiaohui Xie. Diffeomorphic image registration with neural velocity field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1869–1879, 2023.

[15] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):8, 2020.

[16] Monica Hernandez, Matias N Bossa, and Salvador Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*, 85:291–306, 2009.

[17] Jacqueline Le Moigne, Nathan S Netanyahu, and Roger D Eastman. *Image registration for remote sensing*. Cambridge University Press, 2011.

[18] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[19] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. 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.

[20] Marc Modat, Pankaj Daga, M Jorge Cardoso, Sebastien Ourselin, Gerard R Ridgway, and John Ashburner. Parametric non-rigid registration using a stationary velocity field. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 145–150. IEEE, 2012.

[21] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5865–5874, 2021.

[22] Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. Transformers in medical imaging: A survey. *Medical image analysis*, 88:102802, 2023.

[23] Jiacheng Shi, Yuting He, Youyong Kong, Jean-Louis Coatrieux, Huazhong Shu, Guanyu Yang, and Shuo Li. Xmorpher: Full transformer for deformable medical image registration via cross attention. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 217–226. Springer, 2022.

[24] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

[25] Zachary Teed and Jia Deng. Tangent space backpropagation for 3d transformation groups. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10338–10347, 2021.

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[27] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1): S61–S72, 2009.

[28] Hanguang Xiao, Xufeng Xue, Mi Zhu, Xin Jiang, Qingling Xia, Kai Chen, Huanqi Li, Li Long, and Ke Peng. Deep learning-based lung image registration: A review. *Computers in biology and medicine*, 165:107434, 2023.