

Symmetries of the image registration problem and their application to learning based solutions

Hastings Greer

April 2024

1 Thesis Statement

Image registration is a problem where the correct solution has inherent symmetries. By formalizing this observation, and then restricting deep networks that solve the registration problem to have these same symmetries, we can boost performance and regularize, simplify, and accelerate training.

2 Introduction

Image registration is a core tool for medical imaging studies. In my thesis, I seek to improve the performance of learning based approaches on this task through implementing losses and architectures for inverse consistency and equivariance. First, as a foundation, I am developing the notation for registration algorithms. By writing registration algorithms as operators, it becomes ergonomic to prove properties of these algorithms at the blackboard. With this notation, I analyze and exploit symmetries (inverse consistency and spatial equivariance) that the solution to the registration problem should have in order to regularize and accelerate training. Finally, we find that the models we have developed through this approach have a novel emergent ability: they can be trained on composite datasets from various parts of the body and modalities, and perform well on all while even generalizing to unseen organ systems. These registration foundation models are currently achieving high impact, and we seek to keep the wind behind this impact by continuing to improve performance.

3 Notation

Careful development of notation is a fundamental component of progress in image registration, and one that is easily overlooked (and difficult to publish in conferences!)

3.1 Pre-existing Notation

The standard in the field of image registration is to describe loss functions such as similarity loss and regularization loss with typeset equations, but represent the registration algorithm itself using a flowchart. [2] Follows this pattern exactly, although the relatively simple internal structure of their approach means that there is little structure to capture with improved notation. Multistep architectures such as [25, 18] use flowcharts along with prose to clarify ambiguities. The need for new notation comes into the forefront in LapIRN [18] where there is significant. Across these papers, there is the standard that the images are functions from a domain to intensities $I : \Omega \rightarrow \mathcal{R}$ and the transform is a function from the image domain to the image domain $\varphi : \Omega \rightarrow \Omega$.

The canonical unsupervised registration loss in this notation is

$$\mathcal{L}_{\text{sim}} = \text{Sim}(I^{\text{moving}} \circ \varphi, I^{\text{fixed}})$$

(In about half of works in the literature, φ is used to indicate the a map from the moving space to the fixed space, and then the map used in the similarity loss is written φ^{-1} to indicate that the map is used to pull back the moving image features. In these works, ' φ ' is never computed, so φ^{-1} just becomes a long name for the pullback map. I have found that it is better to always talk about the pullback map from the fixed space to the moving space and denote it with a single Greek letter, as otherwise the equations always have \square^{-1} dotted about in an unhelpful way, especially as writing it on every map slows whiteboard analysis, I want to use the superscript space for other information, and only including the inverse symbol sometimes is tremendously confusing.)

4 Registration algorithms, not maps, as the primary object of study

My central concept in notation is to describe registration algorithms $\Phi : ((\Omega \rightarrow \mathcal{R}) \times (\Omega \rightarrow \mathcal{R})) \rightarrow (\Omega \rightarrow \Omega)$.¹ These algorithms take in a pair of images, and return a map. Written with this notation, the unsupervised registration loss becomes

$$\mathcal{L}_{\text{sim}} = \text{Sim}(I^{\text{moving}} \circ \Phi[I^{\text{moving}}, I^{\text{fixed}}], I^{\text{fixed}})$$

By focusing on the properties of registration algorithms Φ instead of maps φ , it becomes easy to succinctly describe symmetries of these algorithms, and to regularize how the transform varies as images change, instead of just how the transform varies in space.

As an example, one of the simplest regularization losses based on a symmetry is to require that the registration algorithm be symmetric.

¹Cyclemorph [16] uses a similar notation, referring to the maps from moving to fixed and fixed to moving as ϕ_{XY} and ϕ_{YX} , but does not elaborate that arbitrary expressions can be substituted for X and Y

In my notation, the inverse consistency loss can be written succinctly and unambiguously as

$$\mathcal{L}_{inv} = ||\Phi[I^A, I^B] \circ \Phi[I^B, I^A] - id||_2^2$$

4.1 Parallelism of Notation and Code

Φ is a torch module, and we take seriously that it returns a function, not a tensor- it returns a python function that maps tensors of coordinates to tensors of coordinates. We have developed a novel python library based on this principle.

For example, a typical simple definition of a convnet for registration Φ is $\Phi_\theta[I^A, I^B](\vec{x}) := \text{Conv}_\theta[\text{cat}(I^A, I^B)](\vec{x}) + \vec{x}$

In our library, this is a class with the following definition:

```
class FunctionFromVectorField(RegistrationModule):
    """
    Wrap an inner neural network 'net' that returns a tensor of displacements
    [B x N x H x W (x D)], into a RegistrationModule that returns a function that
    transforms a tensor of coordinates
    """

    def __init__(self, net):
        super().__init__()
        self.net = net

    def forward(self, image_A, image_B):
        tensor_of_displacements = self.net(image_A, image_B)
        displacement_field = self.as_function(tensor_of_displacements)

        def transform(coordinates):
            return coordinates + displacement_field(coordinates)

        return transform
```

The direct correspondence between notation in publications and code in our open source libraries accelerates development. This library, developed in the course of this thesis, `icon_registration`, has been downloaded 523 times in the past month from pypi.

5 Literature Review

5.1 Unsupervised image registration.

In addition to prior work on equivariance, we also build on research into unsupervised image registration. The foundational concept [2] driving this subfield is to predict deformations using a neural network. Training minimizes a similarity

and a regularity loss. It is feasible to learn multiple steps of registration by simply penalizing the final similarity and regularity [25, 9, 26], even when the steps have different transform models [10, 25]. Diffeomorphisms can be guaranteed by special vector field integration layers [7], inversion layers [12] or *encouraged* by a loss, such as diffusion [2, 25], or inverse consistency [9, 26]. Accuracy can be improved by learning features, computing cost volumes from the similarity between features in the fixed and moving images, and predicting a transform based on this cost volume: the final image similarity then is sufficient to learn good features [18].

5.2 Distributable registration neural networks

Most current deep registration approaches are intended to be used by deciding on the task to be solved, and then training a model. Voxelmorph [2] has achieved great impact with this model, many other papers have achieved little impact despite improving on Voxelmorph’s performance because experimenting with different registration network

There are several ongoing efforts to develop and distribute a neural network where the default mode of use is to download weights and then use them for your task, without performing training. The central attribute distinguishing this approach is that a paper developing a distributable registration network will evaluate their trained model on a different dataset than they trained it on. This gives an estimate of how it will perform on the customer’s dataset.

There are three main ways that a downstream registration task is likely to differ from the training set: Different organ system, different field of view, different imaging modality.

SynthMorph[?] trained their network on synthetic brain images, and achieved high accuracy for brain images from a variety of datasets. They conquer different imaging modalities by broadening the imaging modalities in the training set to encompass the customer imaging modality. They punt on organ system by focusing on brain registration. They fail to address the issue of different field of view, requiring careful work by the customer to bring their data into the SynthMorph default field of view.

EasyReg[?] added common sense registration pipeline components to the SynthMorph model: built in brain extraction, robust affine pre-registration to an atlas coordinate system, built in intensity pre-processing. These serve to bring arbitrary brain image *in distribution* to the existing SynthMorph model, allowing plug and play deformable brain registration. The result is an extremely practical whole brain registration product.

BrainMorph[?] takes a different tack to the field of view issue. By using the equivariance trick from KeyMorph[?], their method’s performance is independent of field of view. They approach the modality question largely by training on diverse modalities.

[?] and [?] tackle the harder challenge of generalizing to arbitrary organ systems. The approach is to simply train on multiple organ systems. They also use this approach from field of view differences, using synthetic augmenta-

tion. MultiGradICON extends this to multiple modalities, both by incorporating many modalities in the training process and synthetically augmenting them. Finally, they close any train-test divergence using instance optimization.

5.3 Inverse Consistent deep registration methods

Inverse consistency in deep image registration approaches is commonly promoted via a penalty [9, 25, 32, 21] on the inverse consistency error. Extensive work also exists on optimization-based *exactly* inverse consistent image registration. For example, by using a symmetric image similarity measure and an inverse consistency loss on the transformations [5] or by performing robust inverse consistent rigid registrations with respect to a middle space [24]. ANTs SyN [1] is an approach to inverse consistent deformable registration, but by default is part of a multi-step affine then SyN pipeline which is not as a whole inverse consistent.

Mok *et al.* [19] introduce a deep-learning framework that is exactly inverse consistent. They take advantage of the fact that a stationary velocity field (SVF) transform representation allows for fast inversion of a transform by integrating the negated velocity field. Thus, by calling their network twice, the second time with the inputs reversed, they can construct a transform $\Phi^{AB} = \exp(N_\theta[I^A, I^B]) \circ \exp(-N_\theta[I^B, I^A])$. This registration network is inverse-consistent by construction, but only supports one step. *Our approach will provide a general inverse consistent multi-step framework.*

Iglesias *et al.* [13] introduce a two-step deep registration framework for brain registration that is inverse consistent by construction. First, they independently segment each image with a U-Net into 97 anatomical regions. The centroids of these regions and the corresponding regions of an atlas are then used to obtain an affine transformation to the atlas. This is inverse consistent. Second, each brain image is resampled to the atlas space followed by an SVF-based transformation, where the velocity field is obtained by two calls to their velocity field network: $\exp(N_\theta[I^A, I^B] - N_\theta[I^B, I^A])$. This symmetrization retains inverse consistency and is conceptually similar to our approach. *However, their approach, unlike ours, does not directly extend to N steps and is not trained end to end.*

5.4 Equivariant Deep Registration Methods

We review related work at the intersection of equivariance (and how to achieve it) and image registration, as well as previous work on unsupervised learning of neural networks for image registration.

Equivariant encoders. The most impactful equivariant layers in the field of neural networks have been ordinary convolution [8] and attention [23], which are equivariant to translation and permutation, respectively. In this work, we focus on using attention and convolution to make a registration algorithm that is by construction $[W, U]$ equivariant to translation. However, there are also more recent approaches to create components with additional geometric equivariances

such as to rotation. These include group equivariant convolution [6], StyleGAN’s equivariant convolution [15], and spherical harmonic convolution [20]. These should be explored in future work to expand the deformations with respect to which our approach (based on what we call *coordinate attention*, see Sec. ??) is $[W, U]$ equivariant.

Keypoint-based equivariant registration architectures. Several registration approaches relying on keypoint matching have been proposed [17, 3]. Notably, any approach that aligns keypoints to keypoints using a least squares fit is $[W, U]$ equivariant to rigid motions if the keypoint selectors and descriptors are equivariant to rigid motions. The reason can be derived from the notion of equivariance as follows: *if two points correspond before warping the input images, they should correspond afterwards*. The keypoints align with each other before and after a warp is applied, and move with features on the input images when they are warped. SURF [3] and related methods fall under this category in the subfield of non-learning-based 2D image alignment. For medical image registration, learning-based approaches that leverage keypoints include EasyReg [14], Keymorph [31], SAME++ [28] and E-CNN [4]. SAME++ [28] uses a pretrained foundation model as a convolutional encoder, and extracts matched keypoints from it using methods traditionally used in 2D image feature alignment, such as checking for cycle consistency. This alignment is then refined using a trained deep network with overall strong performance on lung, abdomen, and head/neck computed tomography (CT) registration tasks. EasyReg, E-CNN and Keymorph achieve equivariant registration by locating a fixed set of keypoints in each image and aligning matching keypoints. EasyReg’s [14] affine pre-registration is by construction $[W, U]$ equivariant to affine deformations because it works by segmenting the brain images to be registered, and then bringing the centroids of the segmented regions into alignment. In effect, this uses each segmented region as a keypoint. In our experiments (see Sec. ??) we will see that this strategy works well even for out-of-distribution translations, and exhibits strong performance when detailed segmentations are available for training. In the KeyMorph work of [30], equivariance is achieved in an unsupervised setting by learning to predict keypoints in brain images. Unlike SURF and related approaches, KeyMorph avoids the need for a matching step by predicting keypoints in a fixed order. Finally, in the E-CNN work of [4], the authors leverage a rotationally equivariant encoder built using spherical harmonics to predict features, and the centers of mass of each feature channel are then used as keypoints and rigidly aligned in a differentiable manner. The network is trained by warping a single image with two different transforms and then penalizing the difference between the network output and the transform used to generate the pair.

Equivariance losses. Several works aim for equivariance via suitably designed loss terms. In the CoMIR work of [22], for instance, the authors train an image encoder to be equivariant to rotations and deformations using an unsupervised contrastive loss with additional constraints. Furthermore, the concept of W -bipath consistency, introduced in [29], is effectively a penalty to encourage the

$[W, U]$ equivariance to B-Spline transforms. In particular, [29] uses penalties to enforce that $\Phi[I^M, I^F \circ U] = \Phi[I^M, I^F] \circ U$ as well as inverse consistency, which together are sufficient conditions for $[W, U]$ equivariance. While this is undoubtedly useful for training (as it is equivariance enforced by a loss), it is not likely to hold for out-of-distribution data, unlike our approach which is *by construction* $[W, U]$ equivariant.

6 Components of the Thesis already published

6.1 ICON

It is well known that image registration models should be inverse consistent. The first three papers of my thesis have involved exploiting this property. The first publication, ICON [9], proved the claim that a standard inverse consistency loss is sufficient to make an ordinary network that predicts

The first publication of the thesis, [9], was focused on a novel approach to regularizing image registration: optimizing for inverse consistency and then getting regularity as a side effect. This approach only had one hyperparameter, which was aspirationally could allow training on a new dataset with less tuning needed than multigaussian regularizers, which have many hyperparameters to manually adjust. This publication introduced the "Registration Network" notation, including explicitly denoting the input images to the registration network, and the operators TwoStep and Downsample that allow hierarchical multistep registration networks to be constructed.

In this work, we also developed and published the library `icon.registration`, upon which the software for all following papers in the thesis as been build.

6.2 GradICON

With ICON we had a strong suspicion that policing the diffeomorphic status of the transform via the inverse consistency error was an excellent idea, as yielded smooth diffeomorphic maps without penalizing large deformations. The ICON formulation had issues in practice because it behaved differently at different resolutions of the displacement field, and fell apart particularly for high resolutions. Lin Tian and I simultaneously independently discovered that this can be ameliorated by penalizing the gradient of the inverse consistency error instead of the error directly. This lead to a large reduction in transform irregularity and increase in convergence speed. This GradICON regularizer appears to dominate ICON regularization, diffusion regularization, and bending energy regularization for any desired level of transform regularity. We submitted this as a joint first author paper to CVPR 2022.

We achieve state-of-the-art registration performance on a variety of real-world medical image datasets using a single set of hyperparameters and a single non-dataset-specific training protocol.

This paper fulfilled the promise of the ICON paper that an inverse consistency based regularizer could be easier to tune than existing approaches.

6.3 ConstrICON

Inverse consistency is a desirable property for image registration. GradICON produced high quality approximately inverse consistent transforms, but there are still some tasks like atlas building where exactly inverse consistent transforms are needed. In ConstrICON [11], we proposed a simple technique to make a neural registration network inverse consistent by construction, as a consequence of its structure, as long as it parameterizes its output transform by a Lie group. We extended this technique to multi-step neural registration by composing many such networks in a way that preserves inverse consistency. This multi-step approach also allows for inverse-consistent coarse to fine registration. We evaluate our technique on synthetic 2-D data and four 3-D medical image registration tasks and obtain excellent registration accuracy while assuring inverse consistency.

Here, we begin to see the power of registration network notation for formally proving properties of multistep registration algorithms.

7 Ongoing work and plan for the Thesis

7.1 Coordinate Attention with Refinement Layers

A desirable property of image registration algorithms is that a correspondence estimate (e.g., the true oracle correspondence) for an image pair is maintained under deformations of the input images. Formally, the estimator should be equivariant to a desired class of image transformations. In "Coordinate Attention with Refinement Layers", I present careful analyses of the desired equivariance properties in the context of multi-step deep registration networks. Based on these analyses I 1) introduce the notions of $[U, U]$ equivariance (network equivariance to the *same* deformations of the input images) and $[W, U]$ equivariance (where input images can undergo *different* deformations); we 2) show that in a suitable multi-step registration setup it is sufficient for overall $[W, U]$ equivariance if the first step has $[W, U]$ equivariance and all others have $[U, U]$ equivariance; we 3) show that common displacement-predicting networks only exhibit $[U, U]$ equivariance to translations instead of the more powerful $[W, U]$ equivariance; and we 4) show how to achieve multi-step $[W, U]$ equivariance via a coordinate-attention mechanism combined with displacement-predicting refinement layers (CARL). Overall, our approach obtains excellent practical registration performance on several 3D medical image registration tasks and outperforms existing unsupervised approaches for the challenging problem of abdomen registration.

This paper continues the theme of calculating properties of multistep registration networks by manipulating registration network notation.

I have submitted CARL to Neurips 2024 and am awaiting a response. If it is rejected, I plan to resubmit to CVPR. For the resubmission, I will add in the 2nd step of training from the GradICON paper to see if this boosts performance, and increase the number of evaluation pairs on the Abdomen1k dataset to 200.

The CVPR submission deadline is in November, but I intend to make these changes after the October 10th ISBI deadline.

7.2 Pursuit of Rotational equivariance

The ICON paper left open the issue of convergence speed and a rapidly escalating regularization hyperparameter, and solving this led to the successful GradICON result. Likewise, the CARL result leaves open the question of rotational equivariance, and solving this open question is an attractive topic for follow on research. I will attempt three routes towards developing a general unsupervised registration architecture with built in $[W, U]$ equivariance to a class of transforms including rotation.

- The Keymorph architecture has unimpressive performance but full rotation equivariance. By our construction, a hybrid keymorph-with-refinement-layers architecture could be made to have full rotational equivariance with the performance bonus of a multi-step method. This would have the downside of precluding a general uniGradICON like model, as the keypoints would be organ-specific. This would also allow fully factoring out the deformation
- The CARL model should have rotational equivariance if an appropriately invariant feature encoder is used in the coordinate attention block. This architecture was not used in CARL due to training instability, but I will make an attempt to resolve this by initializing a rotationally equivariant CARL model using known transforms, following the Keymorph training procedure.
- There are hints that the processes of instance optimization and iteration of the TwoStep operator are inherently equivariant. These hints will be fleshed out into formal proofs, and if these proofs suggest an architecture, this will be pursued.

I will evaluate these approaches on the task of atlas generation for the OAI dataset, and submit this atlas building approach as a four page paper to ISBI, with deadline October 11th.

7.3 Similarity Measure investigation, and Registration Foundation Models

The final paper that I aim to produce in this thesis is an extension of the sequence uniGradICON and multiGradICON. These two papers, which I helped to coauthor, developed a foundation model based on the GradICON architecture

and training procedure, which is general- the same model is used across different data sets. In the HugeGradICON paper I will scale both the dataset diversity and model size- this is of course what you do with a foundation model. Where the "Equivariance of Iterated Registration Steps" Paper leans heavily toward methodological novelty and mathematical tricks, the HugeGradICON paper will be entirely practical, aiming to boost the performance of the approach on the kinds of cases that it is being used on in the wild. We have a unique opportunity to learn about its real world useage from github issues:

Users will use the registration tool on non-brain-extracted images. The model needs to either enforce brain extraction for brain registration, as is done in EasyReg [?], or perform adequately when brain extraction is not performed. We will augment the training set with non-brain-extracted images, and test performance without brain extraction, and add an open source brain extraction model to the CLI if performance without it is not adequate.

In addition, I will use the label image / input image distinction and randomization strategy introduced in MultiGradICON to attempt to boost registration performance on non-brain-extracted images, by sometimes running the model on the brain extracted image but evaluating the loss on the full head image, and vice versa.

Users do not always have access to original images before resampling. Therefore, our model needs to accurately register images that have black bars introduced by the resampling procedure. To achieve this, I will use the "Label image" / "Input" image distinction again, by only adding black bars to the network input image, but not the image on which the loss is evaluated.

Multimodal registration is necessary. The multiGradICON research has found that adding multimodal registration to the training set does improve performance on multimodal tasks, but harms performance on unimodal tasks. This should be ameliorated by increasing network capacity. I intend to further boost multimodal registration performance by adding SynthMorph shapes and brains datasets to the composite dataset.

Finally, for some tasks rotational equivariance or exact inverse consistency are required, and these properties we will experiment with replacing the GradICON backbone with the ConstrICON and CARL backbones.

All performance will be evaluated on the tasks used for GradICON, uni-GradICON and multiGradICON, allowing easy comparison to current state of the art.

This work will be my focus after the november resubmission of the CARL paper, and I intend to submit HugeGradICON to MICCAI in March 2025.

7.4 Defense Date

I aim to defend my thesis at the end of the spring 2025 semester.

8 Coursework and Graduation Requirements

To graduate, I need to take 915 Professional Writing and 630 Graduate Operating systems. I plan to take both in Spring 2025, as neither is offered in Fall 2024.

References

- [1] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *MedIA*, 12(1):26–41, 2008.
- [2] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. VoxelMorph: a learning framework for deformable medical image registration. *TMI*, 38(8):1788–1800, 2019.
- [3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *ECCV*, 2006.
- [4] Benjamin Billot, Daniel Moyer, Neerav Karani, Malte Hoffmann, Esra Abaci Turk, Ellen Grant, and Polina Golland. Equivariant and denoising CNNs to decouple intensity and spatial features for motion tracking in fetal brain MRI. In *MIDL*, 2023.
- [5] Gary E Christensen and Hans J Johnson. Consistent image registration. *TMI*, 20(7):568–582, 2001.
- [6] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *ICML*, pages 2990–2999. PMLR, 2016.
- [7] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *MICCAI*, 2018.
- [8] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.
- [9] Hastings Greer, Roland Kwitt, François-Xavier Vialard, and Marc Niethammer. ICON: Learning regular maps through inverse consistency. In *ICCV*, 2021.
- [10] Hastings Greer, Lin Tian, Francois-Xavier Vialard, Roland Kwitt, Sylvain Bouix, Raul San Jose Estepar, Richard Rushmore, and Marc Niethammer. Inverse consistency by construction for multistep deep registration. In *MICCAI*, pages 688–698. Springer, 2023.

- [11] Hastings Greer, Lin Tian, François-Xavier Vialard, Roland Kwitt, Sylvain Bouix, Raúl San José Estépar, Richard J. Rushmore, and Marc Niethammer. Inverse consistency by construction for multistep deep registration. *ArXiv*, abs/2305.00087, 2023.
- [12] Joel Honkamaa and Pekka Marttinen. Asymreg: Robust symmetric image registration using anti-symmetric formulation and deformation inversion layers. *arXiv preprint arXiv:2303.10211*, 03 2023.
- [13] Juan Eugenio Iglesias. EasyReg: A ready-to-use deep learning tool for symmetric affine and nonlinear brain MRI registration. 2023.
- [14] Juan Eugenio Iglesias. A ready-to-use machine learning tool for symmetric multi-modality registration of brain mri. *Scientific Reports*, 13(1):6657, 2023.
- [15] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *NeurIPS*, 34:852–863, 2021.
- [16] Boah Kim, Dong Hwan Kim, Seong Ho Park, Jieun Kim, June-Goo Lee, and Jong Chul Ye. CycleMorph: Cycle Consistent Unsupervised Deformable Image Registration. *arXiv e-prints*, page arXiv:2008.05772, August 2020.
- [17] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.
- [18] Tony CW Mok and Albert Chung. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In *MICCAI*, 2020.
- [19] Tony C.W. Mok and Albert C.S. Chung. Fast symmetric diffeomorphic image registration with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [20] Daniel Moyer, Esra Abaci Turk, P Ellen Grant, William M Wells, and Polina Golland. Equivariant filters for efficient tracking in 3d imaging. In *MICCAI*, pages 193–202. Springer, 2021.
- [21] Abdullah Nazib, Clinton Fookes, Olivier Salvado, and Dimitri Perrin. A multiple decoder CNN for inverse consistent 3D image registration. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 904–907, 2021.
- [22] Love Nordling, Johan Öfverstedt, Joakim Lindblad, and Natasa Sladoje. Contrastive learning of equivariant image representations for multimodal deformable registration. *ISBI*, pages 1–5, 2023.

- [23] Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. In Jian Su, Kevin Duh, and Xavier Carreras, editors, *EMNLP*, pages 2249–2255, Austin, Texas, November 2016. Association for Computational Linguistics.
- [24] Martin Reuter, H. Diana Rosas, and Bruce Fischl. Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, 53(4):1181–1196, 2010.
- [25] Zhengyang Shen, Xu Han, Zhenlin Xu, and Marc Niethammer. Networks for joint affine and non-parametric image registration. In *CVPR*, 2019.
- [26] Lin Tian, Hastings Greer, Francois-Xavier Vialard, Roland Kwitt, Raúl San José Estépar, and Marc Niethammer. GradICON: Approximate diffeomorphisms via gradient inverse consistency. *CVPR*, pages 18084–18094, 2022.
- [27] Lin Tian, Hastings Greer, François-Xavier Vialard, Roland Kwitt, Raúl San José Estépar, Richard Jarrett Rushmore, Nikolaos Makris, Sylvain Bouix, and Marc Niethammer. GradICON: Approximate diffeomorphisms via gradient inverse consistency, 2022.
- [28] Lin Tian, Zi Li, Fengze Liu, Xiaoyu Bai, Jia Ge, Le Lu, Marc Niethammer, Xianghua Ye, Ke Yan, and Daikai Jin. SAME++: A self-supervised anatomical embeddings enhanced medical image registration framework using stable sampling and regularized transformation. *arXiv preprint arXiv:2311.14986*, abs/2311.14986, 2023.
- [29] Prune Truong, Martin Danelljan, Fisher Yu, and Luc Van Gool. Warp consistency for unsupervised learning of dense correspondences. In *ICCV*, 2021.
- [30] Alan Q. Wang, Evan M. Yu, Adrian V. Dalca, and Mert Rory Sabuncu. A robust and interpretable deep learning framework for multi-modal registration via keypoints. *MedIA*, 90:102962, 2023.
- [31] Evan M. Yu, Alan Q. Wang, Adrian V. Dalca, and Mert R. Sabuncu. KeyMorph: Robust multi-modal affine registration via unsupervised keypoint detection. In *MIDL*, 2022.
- [32] Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *CoRR*, abs/1809.03443, 2018.