uniGradICON: A Foundation Model for Medical Image Registration

Lin Tian[1], Hastings Greer[1], Roland Kwitt[2], François-Xavier Vialard[3], Raúl San José Estépar[4], Sylvain Bouix[5], Richard Rushmore[6], and Marc Niethammer[1]

[1] University of North Carolina at Chapel Hill
[2] University of Salzburg, Austria
[3] University Paris-Est
[4] Brigham and Women's Hospital
[5] ÉTS Montréal
[6] Boston University

Abstract. Conventional medical image registration approaches directly optimize over the parameters of a transformation model. These approaches have been highly successful and are used generically for registrations of different anatomical regions. Recent deep registration networks are incredibly fast and accurate but are only trained for specific tasks. Hence, they are no longer generic registration approaches. We therefore propose `uniGradICON`, a first step toward a foundation model for registration providing 1) great performance *across* multiple datasets which is not feasible for current learning-based registration methods, 2) zero-shot capabilities for new registration tasks suitable for different acquisitions, anatomical regions, and modalities compared to the training dataset, and 3) a strong initialization for finetuning on out-of-distribution registration tasks. `UniGradICON` unifies the speed and accuracy benefits of learning-based registration algorithms with the generic applicability of conventional non-deep-learning approaches. We extensively trained and evaluated `uniGradICON` on twelve different public datasets. Our code and weight are available at https://github.com/uncbiag/uniGradICON.

Keywords: Medical Image Registration · Foundation Models

1 Introduction

Conventional registration methods [2,19,22,13] directly estimate spatial correspondences for an image pair. They can be used for a wide variety of registration tasks, can be highly accurate, but are often slow as they estimate registration parameters from scratch for every registration pair by numerical optimization. More recent supervised [36,4,30] and unsupervised [10,3] learning-based registration approaches *predict* spatial correspondences much faster using a deep registration network. These learning-based approaches have achieved significant accuracy improvements by advanced transformation models [29,27,27], network

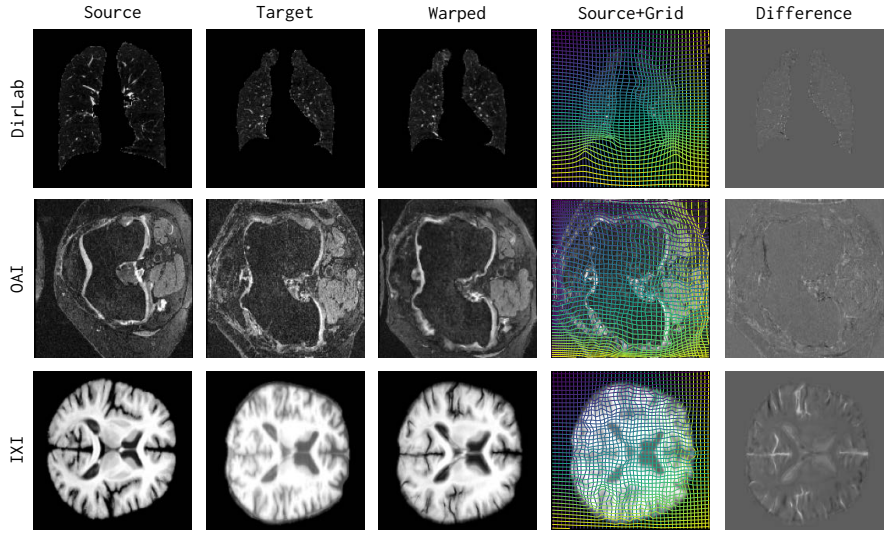

Fig. 1. Example `uniGradICON` registrations. Prediction only w/o IO.

structures [23,6], training schemes [14,9,24], and similarity measures [33,25]. Learning-based methods now outperform conventional methods on various registration tasks in terms of accuracy and speed. However, *current learning-based methods require training task-specific networks* making them much less flexible than approaches that use numerical optimization per image pair.

Our **key question** is: Can we train a *universal* registration network that can be used as generically as conventional registration algorithms while retaining the speed and accuracy advantages of learning-based, but task-specific, methods? One can imagine obtaining such a model by training *one* registration network over many different datasets to obtain a *universal foundation model* for registration. However, this is not straightforward. The crux is that picking the right kind of registration hyperparameters (for regularizer and similarity measure) is important for good registration performance for conventional and learning-based registration alike. However, while this tuning for conventional approaches might be tedious, learning-based approaches generally require the costly training of a new network instance. While some work to adapt hyperparameters after training exists for deep registration networks [16], the fundamental issue is that *different* registration tasks generally require *different* hyperparameters. However, a universal registration network will be trained using only one *fixed* set of hyperparameters, which then would be suboptimal for some registration tasks.

Recent work [12,32] on learning-based registration allows training task-specific registration networks *using exactly the same training procedure and hyperparameters*. This is made possible [32] by replacing conventional regularizers (e.g., diffusion) by gradient inverse consistency (`GradICON`) regularization. While conventional regularizers need to be carefully balanced with an image similarity

measure the GradICON regularizer is weaker as it only encourages invertibility of the transform. This weaker regularity allows the network to discover what transformations are supported by the data and thereby facilitates training with the same hyperparameters across different datasets. Hence, a key question we explore is if GradICON regularization can also be used to train a universal foundation model for registration.

Contributions. 1) We develop (to the best of our knowledge) the first foundation model for registration; 2) we show that our uniGradICON model can achieve excellent registration accuracy across multiple datasets; 3) we demonstrate that uniGradICON can be successfully used to register images from different image sources, anatomical regions, and image modalities.

2 Material & method

2.1 Dataset curation and pre-processing

Dataset. We created a *composite training dataset* from publicly available medical image corpora. This composite dataset contains various anatomical regions (e.g., lung, knee, brain, and abdomen), different modalities (e.g., CT, CBCT, and MRI), and various deformation patterns (e.g., lung inspiration/expiration or inter-subject anatomical mappings). See Tab. 6 in the appendix for details.

Intensity pre-processing. For CT images, we clip the Hounsfield Units (HU) to $[-1000, 1000]$ and then linearly normalize to $[0, 1]$. For MRIs, we clip the maximum intensity at the 99th percentile and then similarily standardize to $[0, 1]$. *This pre-processing is the same across the training and test phases.*

Spacing pre-processing. We resize all images to $[175, 175, 175]$ using trilinear interpolation. Thus, image spacing of the network input images may not be isotropic. We always evaluate on the original images by linearly interpolating the output transformation fields back to the original spacing.

2.2 Registration network

We rely on the publicly-available GradICON registration network [32] which uses a *two-step* registration process: 1) images are run through a three-level multi-resolution registration network where, at each level, a UNet accepts the warped image from the previous level and the target image. In addition, the input images are downsampled to $\frac{1}{4}, \frac{1}{2}, 1$ for each level, respectively; 2) images are run through one UNet that accepts the warped image from the first step and the target image at full resolution. All four UNets have the same architecture.

2.3 Training protocol and experimental setting

Training dataset. Our composite dataset (Tab. 6 in the appendix) for training contains intra- and inter-patient data. The intra-patient dataset (dataset 1) contains 899 pairs of inspiration/expiration lung CT images. The inter-patient

datasets (dataset 2-4) contain 2532, 1076, and 30 images. We randomly sample two images within each dataset, leading to 3,205,512, 578,888, and 450 possible distinct pairs, respectively. To prevent bias due to the differing numbers of paired images for the different datasets, we randomly sample 1000 image pairs from each dataset during each epoch, resulting in 4000 3D image pairs per epoch.
Training loss. The loss proposed in `GradICON` [32] has the following formulation:

$$\mathcal{L} = \mathcal{L}_{\text{sim}}\left(I^A \circ \Phi^{AB}, I^B\right) + \mathcal{L}_{\text{sim}}\left(I^B \circ \Phi^{BA}, I^A\right) + \lambda \|\nabla\left(\Phi^{AB} \circ \Phi^{BA}\right) - \mathbf{I}\|_F^2 \ . \quad (1)$$

Given an ordered image pair (I^A, I^B), the registration network outputs the transformation map Φ^{AB} which maps I^A to the space of I^B. By swapping the input pair (I^B, I^A), we obtain the estimated inverse map Φ^{BA}. The similarity loss $\mathcal{L}_{\text{sim}}$ is computed between the warped image $I^A \circ \Phi^{AB}$ and the target image I^B, and vice versa. We use localized normalized cross correlation $(1 - \text{LNCC})$ as similarity measure. The third term in Eq. (1) is the gradient inverse consistency regularizer, which penalizes differences between the Jacobian of the composition of Φ^{AB} with Φ^{BA} and the identity matrix $\mathbf{I}$; $\|\cdot\|_F^2$ is the Frobenius norm, $\lambda > 0$.

We use `GradICON` as the basic building block of our approach because it provides excellent registration performance for a variety of datasets *using exactly the same hyperparameter and training settings* [32]. We expect this behavior to allow training a better registration model *across* our composite dataset compared to using competing approaches that rely on task-specific training and hyperparameter settings. Sec. 3 empirically supports this hypothesis.
Training hyperparameters. We train the first step for 800 epochs and the second for 200 epochs with a learning rate of 5e-5 and a balancing constant of $\lambda = 1.5$. These are the *default settings* for `GradICON`[7].
Model availability and development plan. We will periodically update our model[8] to include more anatomical regions, modalities, and deformation types.

3 Experiments

`UniGradICON`'s contributions are as follows. **First**, it obtains state-of-the-art (SOTA) or close-to SOTA accuracy *without retraining*, resulting in similar generality as conventional registration approaches. Hence, our model bridges the gap between the versatility of conventional optimization-based registration algorithms (e.g., `ANTs` [2], `Elastix` [19], or `NifyReg` [22], which while largely motivated by brain registration are general purpose registration tools) and the speed and accuracy of task-specific deep registration networks. **Second**, it can provide a satisfying baseline for zero-shot out-of-distribution registrations. **Third**, when combined with finetuning on an out-of-distribution dataset, it can provide on-par performance to task-specific registration networks for that dataset. To

[7] Better performance might be achievable by further hyperparameter tuning. But this is not the focus of the current study which is targeted at establishing if it is feasible to train a deep registration network with good performance across multiple datasets.
[8] https://github.com/uncbiag/uniGradICON

	COPDGene		OAI		HCP		L2R-Abdomen	
	mTRE[mm]	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$
Initial	23.36		7.6		53.4		25.9	
VM-lung [3]	9.88 [32]	0.0	-	-	-	-	-	-
GradICON-lung [32]	1.93 [32]	3e-4	38.0	7.9e-4	73.0	4.4e-5	18.1	3.4e-2
GradICON-lung(IO) [32]	1.31 [32]	2e-4	70.0	8.0e-3	79.3	1.9e-4	36.9	6.6e-1
LapIRN [24]	2.92 [32]	0.0	-	-	-	-	-	
VM+Affine-knee [3]	-	-	66.1 [32]	1.3e-3	-	-	-	-
VM-knee [3]	-	-	46.1 [32]	2.8e-3	-	-	-	-
GradICON-knee [32]	-	-	70.1 [32]	2.61e-2	-	-	-	-
GradICON-knee(IO) [32]	-	-	71.2 [32]	4.20e-3	-	-	-	-
GradICON-brain [32]	-	-	-	-	78.7 [32]	1.2e-3	-	-
GradICON-brain(IO) [32]	-	-	-	-	80.5 [32]	4e-4	-	-
SyN [2]	1.79 [32]	-	65.7 [32]	0	75.8 [32]	0	25.2	0
VoxelMorph-SVF [3]	19.21	0	55.0	1.1e-4	44.2	1.7e-2	33.8	7.2e-2
uniGradICON	2.26	9.3e-5	68.9	3.9e-2	76.2	6.4e-5	48.3	3.1e-1
uniGradICON(IO)	1.40	9.0e-5	70.3	2.2e-2	78.9	2.2e-4	52.2	9.6e-1

Table 1. Comparison between task-specific ( ) and universal ( ) models based on VoxelMorph, LapIRN and uniGradICON. References indicate a published result.

evaluate uniGradICON w.r.t. these three aspects, **we test on 1)** in-distribution test datasets (Sec. 3.1), **2)** out-of-distribution datasets with zero-shot inference (Sec. 3.2), and **3)** by finetuning on an out-of-distribution dataset (Sec. 3.3).

3.1 Performance on in-distribution tasks

We evaluate uniGradICON on datasets 5-7 (Tab. 6) and the validation set of dataset 4 for in-distribution performance. Tab. 1 shows that uniGradICON achieves comparable performance to models trained *specifically* for a dataset. We further observe that uniGradICON 1) generalizes much better across datsets than a task-specific model (GradICON-lung), 2) consistently outperforms an excellent conventional registration approach (SyN), and 3) performs significantly better than a VoxelMorph-based foundation model (Tab. 7 in the appendix), also trained on the composite dataset. These results verify our hypothesis that the weaker GradICON regularizer of uniGradICON indeed allows successfully training *one* universal registration model compared to the lower accuracy of a diffusion-regularizer-based VoxelMorph variant. We did not succeed in training a LapIRN-based [24] foundation model, likely due to the need for task-specific hyperparameter tuning. Also, we note that the universal VoxelMorph model may benefit from affine pre-alignment which is not needed for uniGradICON. However, due to the large number of training samples for a universal model, affine alignment would have to rely on on-the-fly affine registration. This in turn would require a universal affine registration network which does not yet exist.

	Anatom. region	Deformation	Acquisition	Modality
In-distribution	✓	✓	✓	✓
Out-distribution (**Type 1**)	✓	-	✗	✓
Out-distribution (**Type 2**)	✗	✗	-	✓
Out-distribution (**Type 3**)	✓	-	-	✗

Table 2. Types of generalization. ✓and ✗ denote whether corresponding data have been included in the composite training dataset. − denotes a data type that we do not strictly test `uniGradICON`'s generalization capability on.

3.2 Performance on out-of-distribution tasks

We evaluate the zero-shot performance of `uniGradICON` on out-of-distribution datasets. We classify out-of-distribution datasets into three categories: **Type 1** comprises datasets that contain the same anatomical regions as the composite training dataset but originate from different sources (e.g., comparing HCP to IXI); **Type 2** are datasets with unseen anatomical regions not covered by the composite dataset; **Type 3** are datasets of modalities not contained in the composite training dataset. Tab. 2 provides an overview of these different types.

	L2R-NLST Validation		L2R-NLST Test	L2R-OASIS Validation		L2R-OASIS Test	IXI Test	
	mTRE[mm]	$\%\lvert J\rvert_{<0}$	**mTRE**[mm]	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**	**DICE**	$\%\lvert J\rvert_{<0}$
Initial	10.22		11.2	57.18		56	40.6	
`Learn2Reg` [15] Top-1	-	-	1.44	-	-	82	-	-
`Learn2Reg` [15] Top-5	-	-	2.04	-	-	78	-	-
`VoxelMorph` [3]	-	-	-	-	-	-	73.2 [6]	1.522
`TransMorph` [6]	-	-	-	-	-	-	75.4 [6]	1.579
`SyN`	3.04	9.8-1	-	75.6	1.5e-2	-	64.5 [6]	<1e-4
`uniGradICON`	2.07	4.7e-4	-	79.0	8.9e-4	-	70.6	7.4e-3
`uniGradICON`(IO)	1.77	2.0e-2	-	79.6	1.9e-3	-	71.3	1.8e-1

Table 3. Evaluation of `uniGradICON` on **Type 1** out-of-distribution tasks with zero-shot inference and instance optimization (IO).

Different sources (Type 1). We test zero-shot inference of `uniGradICON` on one lung dataset, L2R-NLST, and two brain datasets, L2R-OASIS and IXI. This experiment studies how `uniGradICON` generalizes to images of a modality and of anatomical regions contained in the composite dataset *but* acquired as part of different studies. As we do not have access to the test set of the `Learn2Reg` challenge, we use the validation set for testing. *This is valid because we do not train on the `Learn2Reg` dataset.* As there are no existing foundation models for registration, we compare to `SyN`. To provide context for how the current task-specific models perform on these datasets, we included the results reported in

	COPDGene		OAI		HCP		L2R-Abdomen Validation		L2R-Abdomen Test
	mTRE	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**
Initial	23.36		7.6		53.4		25.9		28
Learn2Reg [15] Top-1	-	-	-	-	-	-	-	-	69
Learn2Reg [15] Top-5	-	-	-	-	-	-	-	-	49
SyN [2]	1.79 [32]	-	65.7 [32]	0	75.8 [32]	0	25.2	0	-
uniGradICON	2.26	9.3e-5	68.9	3.9e-2	76.2	6.4e-5	48.3	3.1e-1	-
uniGradICON(wo Abdomen)	2.20	5.6e-6	68.9	8.9e-3	76.8	1.2e-5	34.1	1.9e-2	-
uniGradICON(wo Abdomen) (IO)	1.41	2.6e-5	70.2	1.2e-2	79.0	1.5e-4	45.3	7.9e-1	-

Table 4. Evaluation of uniGradICON on **Type 2** out-of-distribution tasks with zero-shot inference and instance optimization (IO).

the Learn2Reg official paper and website for the top 5 methods. Tab. 3 shows that uniGradICON performs better than SyN across all three registration tasks with **zero-shot inference**, with further improvements achievable by instance optimization (IO). The uniGradICON results are within the performance range of the top 5 Learn2Reg methods trained and tuned for the specific tasks. Note that while these results are not directly comparable, we assume that the validation and test sets from the same dataset share the same distribution and, hence, share the same trends. We conclude that uniGradICON is a strong out-of-the-box benchmark model for **Type 1** out-of-distribution tasks.

Different regions (Type 2). We test the generalizability of uniGradICON to registrations for unseen anatomical regions. We train uniGradICON excluding L2R-Abdomen from the composite dataset and test on the L2R-Abdomen validation set. This is challenging as the images and deformation patterns are both not seen during training. Tab. 4 shows that although uniGradICON(wo Abdomen) increases the initial DICE score from 25.9% to 34.1%, we observe an accuracy drop compared to uniGradICON which was trained on the full composite dataset. However, uniGradICON(wo Abdomen) achieves a better registration result than SyN. It also provides a good initialization for instance optimization, mitigating most of the performance drop and coming close to the Top-5 Learn2Reg accuracy for a task-specific model. We conclude that uniGradICON can be a good out-of-the-box baseline for **Type 2** out-of-distribution tasks.

Different Modalities (Type 3). We test how uniGradICON generalizes when the input images have different modalities from the composite training dataset. We use the L2R-CBCT and the L2R-MRCT datasets for evaluation. The L2R-CBCT dataset contains paired images of CT and CBCT where both the CBCT and the combination of CT and CBCT are absent in the composite dataset. For the L2R-CTMRI dataset, the input combination of CT and MRI is unseen during training. Tab. 5 shows that although uniGradICON has not been trained for multi-modal registration, its accuracy is within the range of the top 5 well-tuned and task-specific methods on L2R-CBCT, highlighting its strong generalization ability to unseen modalities and multi-modality registration problems. Compared to the excellent performance on L2R-CBCT, uniGradICON is not as

	L2R-CBCT Validation		Test	L2R-CTMR Validation		Test
	DICE	$\%\lvert J\rvert_{<0}$	**DICE**	**DICE**	$\%\lvert J\rvert_{<0}$	**DICE**
Initial	31.3		28.0	31.3		33
`Learn2Reg` [15] Top-1	-	-	63.2	-	-	75
`Learn2Reg` [15] Top-5	-	-	56.9	-	-	71
`SyN` [2]	57.4	0	-	45.0	0	-
`uniGradICON`	57.0	4.7e-4	-	50.0	4e-2	-
`uniGradICON` (IO)	59.9	0	-	66.8	6.1e-1	-
`uniGradICON` (finetune)	60.3	3.6e-1	-	-	-	-
`uniGradICON` (finetune+IO)	63.7	8.9e-1	-	-	-	-

Table 5. Evaluation of `uniGradICON` on **Type 3** out-of-distribution tasks with zero-shot inference, instance optimization (IO), and target task finetuning.

strong as the well-tuned and trained task-specific methods on L2R-CTMRI. We hypothesize that the combination of CBCT and CT is visually closer to the CT pairs in the composite dataset than the combination of MRI and CT. Thus, it is more challenging for `uniGradICON` to generalize to the L2R-CBCT task. We conclude that `uniGradICON` can be used as an out-of-the-box baseline method for **Type 3** out-of-distribution tasks.

3.3 Performance of finetuning on out-of-distribution dataset

We study the performance of `uniGradICON` when used as an *initialization* and finetuned on a target registration task. We test on the **Type 3** out-of-distribution dataset L2R-CBCT. We finetune `uniGradICON` with the L2R-CBCT training dataset (excluding the validation set) for 4,000 epochs with the learning rate and hyper-parameters used initially. Tab. 5 shows that the finetuned `uniGradICON` model is better than the best task-specific `Learn2Reg` model.

4 Conclusion, limitations, and future work

We have developed `uniGradICON`, a foundation registration model that performs on par with task-specific SOTA methods for in-distribution registration tasks (Tab. 1), alleviating the burden of training new registration networks for every task. `UniGradICON` achieves comparable performance to well-trained SOTA task-specific methods on datasets collected from different sources (Tab. 3) and that contain out-of-distribution modalities (Tab. 5), demonstrating `uniGradICON`'s good out-of-the-box baseline registration performance. We also showed that finetuning `uniGradICON` on an unseen target dataset can further improve accuracy.
Limitations and future work. First, more training datasets could be included for the training of `uniGradICON`. Although `uniGradICON` shows multi-modal generalization abilities (cf. Tab. 5), its support for multi-modal registration could be improved by training on multi-modal image datasets, or by using $1 - \text{LNCC}^2$ or

normalized mutual information as the similarity measure. A self-supervised representation may also help improve the current limited zero-shot performance for unseen regions (cf. Tab. 4). UniGradICON only uses images: further improvements might also be possible by including segmentations for training and instance optimization. Finally, we remark that uniGradICON uses *the* GradICON deep network; using a larger network most likely improves performance.

Acknowledgements. This work was supported by NIH grants 1R01AR072013, 1R01AR082684, 1R01EB028283, 1R21MH132982, RF1MH126732, 1R01HL149877, 5R21LM013670, and R01NS125307. The work expresses the views of the authors, not of NIH. Roland Kwitt was supported in part by the Land Salzburg within the EXDIGIT project 20204-WISS/263/6-6022, 0102-F1901166-KZP, and 20204-WISS/225/197-2019. Sylvain Bouix was supported in part by Natural Sciences and Engineering Research Council grants RGPIN-2023-05443 and CRC-2022-00183. The knee imaging data were obtained from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive. Dataset identifier: NIMH Data Archive Collection ID: 2343. The brain imaging data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The lung imaging data was provided by the COPDGene study. Further data was provided by the Learn2Reg challenge as well as through IXI (Information eXtraction from Images – EPSRC GR/S21533/02).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akin, O., Elnajjar, P., Heller, M., Jarosz, R., Erickson, B., Kirk, S., et al.: Radiology data from the cancer genome atlas kidney renal clear cell carcinoma [TCGA-KIRC] collection. The Cancer Imaging Archive (2016)
2. 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)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. TMI **38**(8), 1788–1800 (2019)
4. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D.: Deformable image registration based on similarity-steered CNN regression. In: MICCAI (2017)
5. Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A.M., et al.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Physics in Medicine & Biology **58**(9), 2861 (2013)
6. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. MedIA **82**, 102615 (2022)
7. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al.: The cancer imaging archive (TCIA): Maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (Jul 2013)
8. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al.: The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging **26**, 1045–1057 (2013)
9. 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)
10. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: DLMIA/MICCAI (2017)
11. Erickson, B.J., Kirk, S., Lee, Y., Bathe, O., Kearns, M., Gerdes, C., et al.: The cancer genome atlas liver hepatocellular carcinoma collection (TCGA-LIHC) (2016)
12. Greer, H., Kwitt, R., Vialard, F.X., Niethammer, M.: ICON: Learning regular maps through inverse consistency. In: ICCV (2021)
13. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: MICCAI (2012)
14. Hering, A., van Ginneken, B., Heldmann, S.: mlvirnet: Multilevel variational image registration network. In: MICCAI (2019)
15. Hering, A., Hansen, L., Mok, T.C., Chung, A.C., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al.: Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. TMI **42**(3), 697–712 (2022)
16. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: Hypermorph: Amortized hyperparameter learning for image registration. In: IPMI (2021)
17. Hugo, G.D., Weiss, E., Sleeman, W.C., Balik, S., Keall, P.J., Lu, J., Williamson, J.F.: Data from 4D lung imaging of NSCLC patients (2016)
18. Hugo, G.D., Weiss, E., Sleeman, W.C., Balik, S., Keall, P.J., Lu, J., et al.: A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer. Medical Physics **44**(2), 762–771 (Feb 2017)

19. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. TMI **29**(1), 196–205 (2009)
20. Linehan, M., Gautam, R., Kirk, S., Lee, Y., Roche, C., Bonaccio, E., et al.: The cancer genome atlas cervical kidney renal papillary cell carcinoma collection (TCGA-KIRP) (2016)
21. 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)
22. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., et al.: Fast free-form deformation using graphics processing units. Computer methods and programs in biomedicine **98**(3), 278–284 (2010)
23. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: CVPR (2020)
24. Mok, T.C., Chung, A.C.: Large deformation diffeomorphic image registration with laplacian pyramid networks. In: MICCAI (2020)
25. Mok, T.C., Li, Z., Bai, Y., Zhang, J., Liu, W., Zhou, Y.J., et al.: Modality-agnostic structural image representation learning for deformable multi-modality medical image registration. arXiv:2402.18933 (2024)
26. Nevitt, M., Felson, D., Lester, G.: The osteoarthritis initiative. Protocol for the cohort study **1**, 737 (2006)
27. Niethammer, M., Kwitt, R., Vialard, F.X.: Metric learning for image registration. In: CVPR (2019)
28. Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., et al.: Genetic epidemiology of COPD (COPDGene) study design. COPD: Journal of Chronic Obstructive Pulmonary Disease **7**(1), 32–43 (2011)
29. Shen, Z., Han, X., Xu, Z., Niethammer, M.: Networks for joint affine and non-parametric image registration. In: CVPR (2019)
30. 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 (2017)
31. Team, T.N.L.S.T.R.: Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine **365**(5), 395–409 (Aug 2011)
32. Tian, L., Greer, H., Vialard, F.X., Kwitt, R., Estépar, R.S.J., Rushmore, R.J., Makris, N., Bouix, S., Niethammer, M.: GradICON: Approximate diffeomorphisms via gradient inverse consistency. In: CVPR (2023)
33. Tian, L., Li, Z., Liu, F., Bai, X., Ge, J., Lu, L., Niethammer, M., Ye, X., Yan, K., Jin, D.: Same++: A self-supervised anatomical embeddings enhanced medical image registration framework using stable sampling and regularized transformation. arXiv:2311.14986 (2023)
34. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., et al.: The Human Connectome Project: a data acquisition perspective. Neuroimage **62**(4), 2222–2231 (2012)
35. Xu, Z., Lee, C.P., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., et al.: Evaluation of six registration methods for the human abdomen on clinically acquired CT. TBE **63**(8), 1563–1572 (2016)
36. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration–a deep learning approach. NeuroImage **158**, 378–396 (2017)