Deformable Image Registration with Geometry-informed Implicit Neural Representations

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

Deformable image registration is a crucial component in the analysis of motion in time series. In medical data, the deformation fields are often predictable to a certain degree: the muscles and other tissues causing the motion-of-interest form shapes that may be used as a geometric prior. Using an Implicit Neural Representation to parameterize a deformation field allows the coordinate space to be chosen arbitrarily. We propose to curve this coordinate space around anatomical structures that influence the motion in our time series, yielding a space that is aligned with the expected motion. The geometric information is therefore explicitly encoded into the neural representation, reducing the complexity of the optimized deformation function. We design and evaluate this concept using an abdominal 3D cine-MRI dataset, where the motion of interest is bowel motility. We align the coordinate system of the neural representations with automatically extracted centerlines of the small intestine. We show that explicitly encoding the intestine geometry in the neural representations improves registration accuracy for bowel loops with active motility when compared to registration using neural representations in the original coordinate system. This approach may improve the efficiency of deformable registration when describing motion-of-interest that is influenced by the shape of anatomical structures.

Keywords: Deformable image registration, geometric prior, cine-MRI, small intestine.

1. Introduction

Deformable registration is the practice of transforming multiple images into a shared coordinate space while minimizing the local correspondences of image content. This is an important step in various medical image processing pipelines. It is used for tasks such as joint analysis of multiple modalities (Wells III et al., 1996), longitudinal disease monitoring (Castadot et al., 2010) and motion estimation in cinematic modalities (Wang and Amini, 2011). Deep learning has been widely applied for deformable image registration in recent years, but it is has not supplanted classical optimization-based methods: such methods still typically outperform novel learning-based methods in public benchmarks and challenges (Hering et al., 2022). Recent work has proposed the use of Implicit Neural Representations (INRs) for deformable image registration (Wolterink et al., 2022), where the deformation function is parameterized with a lightweight multi-layer perceptron that can be optimized during test-time. This approach outperformed various learning-based methods on the DIR-Lab benchmark.

In the paradigm of registration with INRs, the optimization target is an estimated analytic function that maps $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. The complexity of this function determines the required

© 2023 Anonymous.

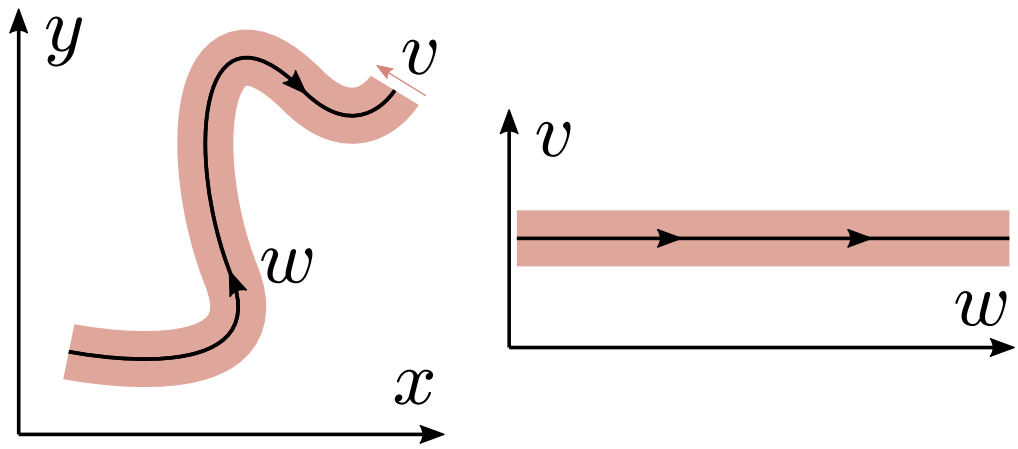

Figure 1: Schematic visualization of the transformation from image coordinate space xyz to tangent space uvw. In intestinal motility images, the dominant motion consists of contractions that are aligned with the intestines. Expressing the motion field in terms of tangent coordinates is likely to result in a simpler analytic function than the same motion field expressed in terms of image coordinates, as the latter would need to implicitly encode the curve w.

capacity of the neural network and the required computational effort to optimize it, meaning the complexity of this function should be minimized. In the general case with arbitrary sets of input images, it is likely that there is no simpler formulation of the deformation function than the one proposed by (Wolterink et al., 2022), mapping image coordinates to deformation vectors. However, in medical motion analysis tasks, the deformation fields are often predictable to a certain degree: the muscles and other tissues causing and constraining the motion-of-interest form shapes that may be used as a geometric prior. The same is true for various disease development tracking tasks, provided the images from both time-points are aligned to a canonical space. We seek to simplify the optimized deformation functions by exploiting such knowledge of the images to be registered.

Geometric deep learning methods have been proposed to exploit symmetries in data (Bekkers et al., 2018; Weiler et al., 2018). By explicitly encoding these symmetries in a non-parametric fashion, the complexity of an approximated function is reduced. This concept has been extended to gauge-equivariant networks to retain the equivariant structure of convolutions on arbitrary surfaces (Cohen et al., 2019). In this work, we aim to achieve a similar effect in registration of intestinal motility images by aligning our coordinate system with the dominant motion. Rather than embedding geometric structure in our network architecture, we condition our coordinate system on the centerline curve of the small intestine. We then optimize the deformation function in a tangent space to this curve. A schematic visualization of this transformation is shown in Figure 1. This approach was inspired by work on dynamical system modelling, where aligning local coordinate frames to

the motion vectors of nodes in the system has been shown to improve performance (Kofinas et al., 2021).

We compare our proposed method with INR-based registration in image coordinates and analyze the differences in the optimization process and registration performance. We evaluate on an intestinal motility dataset that contains small intestine segments from healthy volunteers, as well as data from inflammatory bowel disease (IBD) patients. A relevant symptom of IBD is reduced intestinal motility (Menys et al., 2018). While some bowel loops in the IBD set express normal motile motion, the average motility in these scans is much lower. Hence, a smaller component of the motion in these cases is aligned with the proposed geometric prior.

2. Data

We use an abdominal 4D MRI dataset that contains scans of 14 healthy volunteers, and 10 patients with IBD that were scheduled for ileocecal resection surgery. Each scan consists of volumetric sequences acquired at 1.0 volume per second during a breath-hold. At least 16 timepoints are available for each case. Volumes were acquired at a resolution of 2.5x2.5x2.5 mm and reconstructed to 1.4x1.4x2.5 mm with an FOV of 400x400x35 mm. Small intestine centerline annotations were automatically generated using the method described by (van Harten et al., 2022). This method yields centerline segments using a neural tracker that terminates either at the border of the field of view, or at the point where the uncertainty of the tracker reaches a threshold. The extracted segments may partially overlap due to tracking mistakes; in cases where multiple extracted centerline segments overlap more than 33%, the shortest segment was discarded. Segments with a length of less than 10 cm were excluded from the analysis, to filter out segments for which the tracking method terminated early. The resulting set contained 117 centerline segments: 67 segments from healthy volunteers and 50 segments from IBD patients.

3. Methods

We propose a method for deformable image registration using implicit neural representations conditioned on small intestine centerlines. We explicitly encode the centerline information into the model by transforming the coordinate space of the images into a tangent frame that is aligned with the centerline curve.

3.1. Implicit neural representations

Implicit Neural Representations are neural networks applied as function approximators, which operate on continuous coordinates rather than image values. The image information only enters the network through backpropagation of gradients from a loss function that relates the coordinates to pixel values.

The method uses neural representation $\Phi(\bar{x}) = u(\bar{x}) + \bar{x}$ mapping coordinates from source domain S to target domain T, where $u(\bar{x})$ is a sinusoidal representation network (SIREN) that parameterizes the deformation vector field. In contrast to previous work (Wolterink et al., 2022), $\bar{x}$ is not sampled from image space $\mathbb{R}^3$, but from tangent space uvw, where w is the curve that describes the small intestine centerline and u, v are basis vectors such

that uvw forms a locally Euclidean tangent space around this centerline. Our optimization objective is:

$$
\begin{aligned}
L &= L^{data} + \alpha L^{jac}, \\
L^{data} &= \frac{1}{bs} \sum_{i=1}^{bs} -NCC(\ T[R(\bar{x}_i)],\ S[R(\Phi(\bar{x}_i))]\ ), \\
L^{jac} &= \frac{1}{bs} \sum_{i=1}^{bs} |1 - \det \nabla \Phi[\bar{x}_i]|,
\end{aligned}
\tag{1}
$$

where α is a weighting factor, bs is the batch size, NCC is the normalized cross correlation and R is the function that maps tangent coordinates to images coordinates.

3.2. Tangent space definition

Tangent coordinate space uvw is constructed as a parallel transport frame (Hanson and Ma, 1995) around the centerline. Given a smooth centerline, this yields a smoothly varying function of rotation matrices along the centerline curve. To ensure a locally Euclidean space around central axis w, we resample each centerline segment as a least-squares polynomial before curve framing. While locally Euclidean, this frame is not an injective function that maps image coordinates to unique coordinates in tangent space: when mapping image coordinates to tangent coordinates, we select the rotation matrix that corresponds to the closest point on the centerline curve.

4. Experiments and results

4.1. Optimization details

The networks for the INRs are SIRENs that contain 3 hidden layers of 256 neurons, initialized with $\omega = 64$. The networks are optimized for $2,500$ iterations using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of $1e{-}4$, a batch size of $10,000$ and regularisation weighting factor $\alpha = 0.05$. The foreground mask is defined as a tube around the intestinal centerline with a diameter of 40 mm. This is twice a typical non-contracted small intestine diameter, chosen to account for possible inaccuracies in the extracted centerlines, as well as for possible pathological distension. We center the foreground mask in the coordinate system and we scale the axes to range $[-1, 1]$, constraining the input domain of the function to a single period of the sinusoidal network activation. For experiments where we compare our proposed tangent space registration with an image space registration baseline, we use the same hyperparameters and optimization procedure as described above. The code is publicly available[1].

4.2. Experiments

We perform pairwise registration for each bowel segment in the dataset, registering the first timepoint to each subsequent timepoint in the sequence. As optimization is a stochastic

1. https://github.com/withheld/withheld

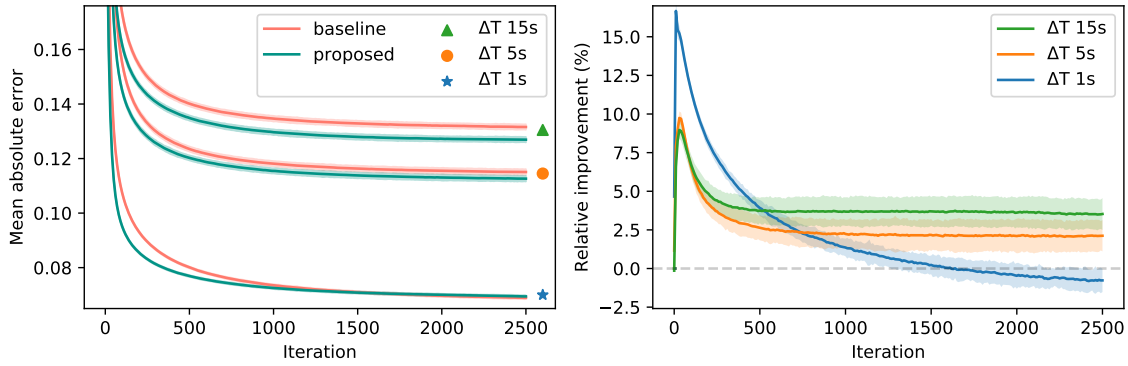

Figure 2: A comparison of the image space registration method and the proposed tangent space registration method in the set of healthy volunteers, averaged over all intestine segments for increasing time differences between images. Left: MAE within the foreground mask after each optimization iteration. Right: The average relative improvement in MAE of the proposed tangent space method compared to the image space baseline. Shaded areas indicate 95% confidence interval.

process, we repeat our experiments with 40 different random seeds to evaluate consistency. To investigate the impact of our centerline prior on the optimization process, we assess the mean absolute error (MAE) between the target image and the transformed source image within the foreground mask after each iteration, for the following experimental settings. First, we compare the optimization performance of the image space registration with the performance of the proposed tangent space registration method in the set of healthy volunteers, for varying differences between the acquisition times of the registered volumes. These results are shown in Figure 2. We observe that for all time differences, the MAE in the first 250 iterations is substantially lower for the proposed tangent space registration method. For small time differences, the registration in image space eventually catches up with the proposed method, whereas for larger time differences, this does not happen.

Several qualitative results are shown in Figure 3. We distinguish three groups of results. The first are cases where the proposed registration with geometric prior is beneficial, as shown in 3.A. In these cases, the dominant motion in the bowel loop is caused by its motility, resulting in a deformation function that is easier to express in tangent coordinates. The second are cases where the proposed method is not beneficial, either due to an incorrectly extracted centerline or due to an absence of motility, as shown in 3.B. The third are cases where aligning the coordinate system with the intestinal centerline is actively harmful to the registration result. This includes cases where the subject is unable to maintain the breath-hold, such that breathing becomes the dominant motion (as shown in 3.C).

Figure 4 shows the results for both the healthy volunteers and the IBD patients, comparing the proposed tangent space registration method with the image space registration baseline for timepoints with a 15 second time difference. These results show that the average relative performance of registration in tangent space coordinates is significantly worse for

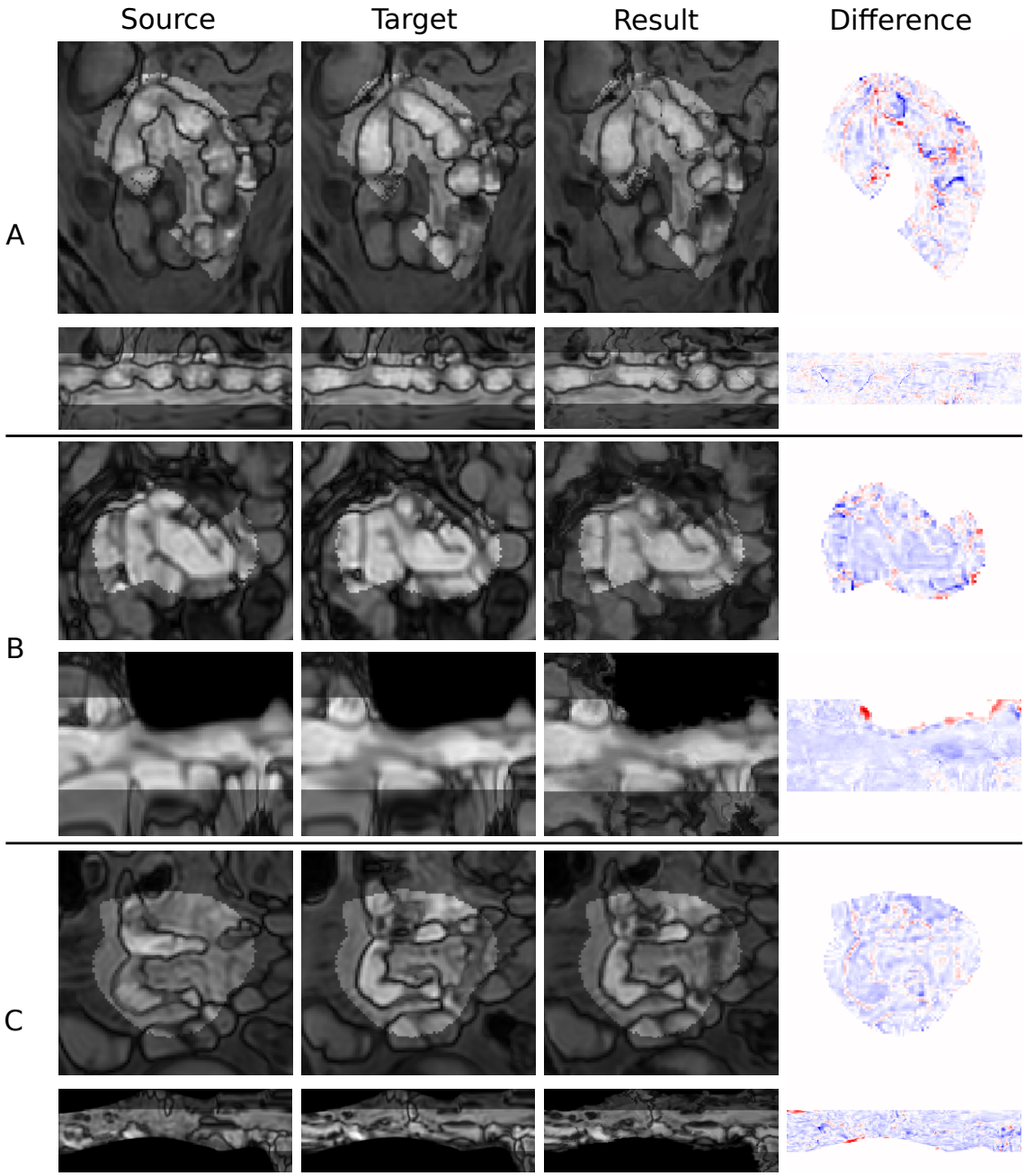

Figure 3: Three examples of registration results from the image space registration (top) and the proposed tangent space registration (bottom). A) Registration with geometric prior is beneficial (16% MAE improvement); the centerline is correctly positioned in both timepoints and there is motile activity. B) The prior is not beneficial (–0.4% MAE improvement); no motile activity is present in the bowel loop. C) The prior is harmful (–42% MAE improvement); breathing motion moved the bowel loop away from the centerline curve in one of the timepoints.

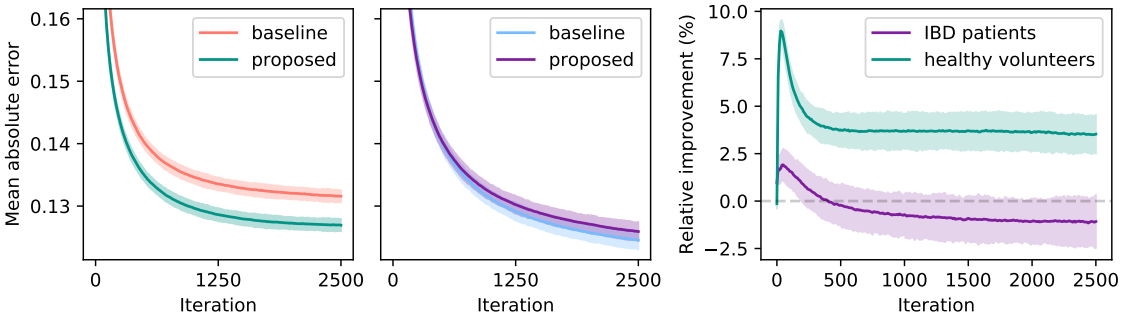

Figure 4: Registration results for a time difference of 15 seconds between source and target images. Shown are the mean absolute error averaged over all centerline segments, evaluated at each iteration in the set of healthy volunteers (left) and IBD patients (middle), as well as the average relative improvement in mean absolute error resulting from the proposed tangent space registration method (right). Shaded areas indicate the 95% confidence interval.

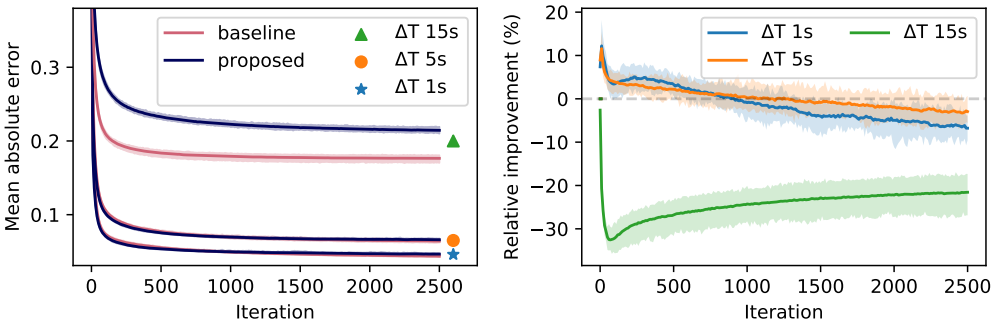

Figure 5: Registration results for a case with breathing motion. Left: The MAE averaged over the centerline segments from one patient for three time differences. Right: The average relative improvement in MAE resulting from the proposed tangent space registration method. Shaded areas indicate 95% confidence intervals.

the set of IBD patients compared to the set of healthy volunteers. Qualitatively inspection of the results reveals this this discrepancy is caused by a combination of fewer segments with active motility, and a number of cases in which breathing motion is present.

The results for the patient with the most severe breathing artifacts are shown separately in Figure 5. For this patient, the registration results from the proposed method are very similar to the results from the baseline method for time differences of 1 and 5 seconds. However, the results are approximately 20% worse than the baseline when the method is applied to the timepoint in which the patient had lost control of their breath.

5. Discussion and Conclusion

We have presented a method for deformable image registration with implicit neural representations that incorporate knowledge of the physical geometry in the optimization process. By aligning the coordinate system of our problem space with small intestine centerline curves, we construct a locally Euclidean tangent space to simplify the estimated function that describes the intestinal motility. In all of our experiments, the proposed tangent space registration method resulted in the average mean absolute error decreasing faster than the image space baseline. This suggests that the functions estimated in tangent space are indeed simpler to optimize than their counterparts in the original coordinate system.

Our proposed tangent space registration method outperforms the image space baseline in bowel segments where motility is present, but has no benefit in dysmotile bowel segments. This is in line with intuition: the geometric information simplifies the function describing the local motion, which is aligned with the centerline curve. Motion from other sources, such as nearby bowel segments, is not aligned with this same curve. In cases where breathing motion is present, the proposed tangent space registration resulted in a performance reduction. Breathing motion can be captured in a simple mathematical function when expressed in image coordinates, as all voxels of interest move in approximately the same direction. Conversely, in tangent coordinates, this function has to implicitly encode the centerline curve to capture the same motion.

In registration of timepoints with small time differences, the proposed method attains adequate registration results faster than the baseline. However, these differences disappear as the models are optimized for more iterations. We assume that this happens because the registration problem for small time differences is sufficiently simple that explicitly providing the shape of centerline curve w has no benefit over implicitly learning this same curve.

The tangent space used in this work is common to the source and target images. It is constructed from a centerline that was automatically extracted in the first timepoint. This relies on the assumption that the abdominal centerlines have little to no movement in position and shape over the course of the breath-hold. In many of our images, this assumption is violated to some degree. Centerlines drift away from their original positions due to the natural movement of the bowels, as well as due to clenching muscles when subjects have trouble holding their breath. It may be possible to reduce the impact of the issues raised above by using rigid alignment before transforming to the tangent coordinate space. This would essentially filter out the rigid components of the deformation function that are simple to express in image space coordinates, but complicated to express in tangent coordinates.

Reducing the complexity of a deformation function by expressing the problem in a more suitable coordinate system is not constrained to parallel transport frames around centerlines. For example, a task where this same concept could be applied is the description of cardiac motion, where a coordinate system constructed based on the heart segmentation may be beneficial for registration performance.

In conclusion, this work has presented a deformable registration method that uses geometry-aware implicit neural representations. These representations are explicitly conditioned on the shapes of anatomical structures that influence the expected motion. Our experiments show that this method may improve the efficiency of deformable registration when describing motion-of-interest that is influenced by known anatomical shapes.

References

Erik J Bekkers, Maxime W Lafarge, Mitko Veta, Koen AJ Eppenhof, Josien PW Pluim, and Remco Duits. Roto-translation covariant convolutional networks for medical image analysis. In *International conference on medical image computing and computer-assisted intervention*, pages 440–448. Springer, 2018.

Pierre Castadot, Xavier Geets, John Aldo Lee, Nicolas Christian, and Vincent Grégoire. Assessment by a deformable registration method of the volumetric and positional changes of target volumes and organs at risk in pharyngo-laryngeal tumors treated with concomitant chemo-radiation. *Radiotherapy and Oncology*, 95(2):209–217, 2010.

Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *International conference on Machine learning*, pages 1321–1330. PMLR, 2019.

Andrew J Hanson and Hui Ma. Parallel transport approach to curve framing. *Indiana University, Techreports-TR425*, 11:3–7, 1995.

Alessa Hering, Lasse Hansen, Tony CW Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Miltiadis Kofinas, Naveen Nagaraja, and Efstratios Gavves. Roto-translated local coordinate frames for interacting dynamical systems. *Advances in Neural Information Processing Systems*, 34:6417–6429, 2021.

Alex Menys, Carl Puylaert, Charlotte E Tutein Nolthenius, Andrew A Plumb, Jesica Makanyanga, Jeroen A Tielbeek, Doug Pendse, Lodewijk A Brosens, Manuel Rodriguez-Justo, David Atkinson, et al. Quantified terminal ileal motility during MR enterography as a biomarker of Crohn disease activity: prospective multi-institution study. *Radiology*, 289(2):428–435, 2018.

Louis D van Harten, Catharina S de Jonge, Kim J Beek, Jaap Stoker, and Ivana Išgum. Untangling and segmenting the small intestine in 3D cine-MRI using deep learning. *Medical Image Analysis*, 78:102386, 2022.

Hui Wang and Amir A Amini. Cardiac motion and deformation recovery from MRI: a review. *IEEE transactions on medical imaging*, 31(2):487–503, 2011.

Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2018.

William M Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.

Jelmer M Wolterink, Jesse C Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *Medical Imaging with Deep Learning*, 2022.