
GradICON: Approximate Diffeomorphisms via Gradient Inverse Consistency

Lin Tian*
UNC Chapel Hill
lintian@cs.unc.edu

Hastings Greer*
UNC Chapel Hill
tgreer@cs.unc.edu

François-Xavier Vialard
LIGM, Université Gustave Eiffel
MOKAPLAN, INRIA Paris
francois-xavier.vialard@u-pem.fr

Roland Kwitt
University of Salzburg
Roland.Kwitt@sbg.ac.at

Raúl San José Estépar
Harvard Medical School
rsanjose@bwh.harvard.edu

Marc Niethammer
UNC Chapel Hill
mn@cs.unc.edu

Abstract

Many registration approaches exist with early work focusing on optimization-based approaches for image pairs. Recent work focuses on deep registration networks to predict spatial transformations. In both cases, commonly used non-parametric registration models, which estimate transformation functions instead of low-dimensional transformation parameters, require choosing a suitable regularizer (to encourage smooth transformations) and its parameters. This makes models difficult to tune and restricts deformations to the deformation space permissible by the chosen regularizer. While deep-learning models for optical flow exist that do not regularize transformations and instead entirely rely on the data these might not yield diffeomorphic transformations which are desirable for medical image registration. In this work, we therefore develop GradICON building upon the unsupervised ICON deep-learning registration approach, which only uses inverse-consistency for regularization. However, in contrast to ICON, we prove and empirically verify that using a *gradient* inverse-consistency loss not only significantly improves convergence, but also results in a similar implicit regularization of the resulting transformation map. Synthetic experiments and experiments on magnetic resonance (MR) knee images and computed tomography (CT) lung images show the excellent performance of GradICON. We achieve state-of-the-art (SOTA) accuracy while retaining a simple registration formulation, which is practically important.

1 Introduction

Image registration is a key component in medical image analysis to estimate spatial correspondences between image pairs [11, 31]. Applications are manifold and include estimating organ motion between treatment fractions in radiation therapy, capturing disease progression, or allowing for localized analyses in a common coordinate system.

Many different registration algorithms have been proposed over the last several decades [24, 23, 37, 38, 6]. Contributions have focused on different transformation models (i.e., what types of transformations are considered permissible), similarity measures (i.e., how "good alignment" between image pairs is quantified), and solution strategies (i.e., how transformation parameters are numerically estimated). The respective choices are generally based on application requirements and assumptions about image appearance and the expected transformation space. In consequence, while reliable registration

*Equal Contribution.

algorithms have been developed for transformation models ranging from simple parametric models (e.g., rigid and affine transformations) to significantly more complex non-parametric formulations [24, 23, 37], which allow highly localized control, practical applications of registration often require many choices and rely on significant parameter tuning efforts to achieve good performance.

Recent image registration work has shifted the focus from solutions based on numerical optimization for a specific image pair to learning to predict transformations based on large populations of image pairs via deep learning [6]. However, while numerical optimization is now replaced by training a deep regression model which can be used to quickly predict transformations at test time, parameter tuning remains a challenge as loss terms for these two types of approaches are highly related (and frequently the same); now one also has to contend with additional choices regarding deep network architectures. Hence, a simple to use deep registration approach, which can adapt to different types of data has remained elusive, in particular, for non-parametric registration approaches, which often require a difficult tuning balance between image similarity and the regularizer of a transformation model to assure good matching with a high level of spatial regularity. This difficulty is compounded in a multi-scale setting where registrations at multiple scales are used to avoid poor local solutions.

Instead of relying on a complex spatial regularizer, the ICON approach by Greer et al. [14] explored using only inverse consistency to regularize the sought-after transformation map thereby dramatically reducing the number of parameters to tune. While inverse consistency is not a new concept in image registration and has been explored to obtain transformations which are inverses of each other when swapping the source and the target images [7], the ICON work showed that inverse consistency results in spatial regularity within a deep registration network setting. Further, as ICON does not explicitly penalize spatial gradients of the deformation field it does not require pre-registration (such as rigid or affine registrations) unlike many other registration approaches. While attractive, the ICON approach suffers from the following limitations: 1) Training convergence is slow, making an ICON model costly to train; 2) Enforcing approximate inverse consistency strictly enough to prevent folds becomes increasingly difficult at higher spatial resolution, necessitating a scheduler for the inverse consistency penalty and making end-to-end training with ICON impractical.

Our GradICON approach is based on a surprisingly simple, initially somewhat counter intuitive, but remarkably effective observation: penalizing the Jacobian of the inverse consistency condition instead of inverse consistency directly² applies zero penalty for inverse consistent transform pairs but yields 1) significantly improved convergence, 2) no longer requires a scheduler for the inverse consistency penalty, and hence 3) allows for end-to-end training. This end-to-end training capability also nicely allows multi-scale training while only specifying the similarity measure and regularizing the transformation at the highest spatial resolution. This is in contrast to existing multi-scale deep registration networks which are trained scale-by-scale and hence often require scale-dependent tuning of the regularizer and the similarity measure. Note that computing a similarity measure at multiple resolutions in a typical multi-scale strategy is desirable as it is a way to discourage poor local optima. GradICON can achieve this as well by using a multi-scale image similarity measure, applied *after* the transformation instead of attempting to register (as of yet unaligned) images at different scales.

Our contributions are as follows:

- We develop GradICON, a versatile deep registration approach which relies on penalizing the Jacobian of the inverse consistency condition and thereby results empirically and theoretically in spatially well regularized transformation maps.
- We show the good performance of GradICON on two large medical datasets: a knee magnetic resonance image (MRI) dataset of the Osteoarthritis Initiative (OAI) and a computed tomography (CT) inhale/exhale lung dataset from COPDGene. GradICON outperforms the SOTA for the OAI dataset and results in competitive performance on COPDGene.
- For the challenging COPDGene images we propose a self-adaptive multi-scale similarity measure, which mimics a multi-scale registration approach, but does so *after* spatial transformation, and is therefore well suited for the alignment of thin vessels and airways.

²I.e., penalizing deviations from $D(\Phi_{\theta}^{AB} \circ \Phi_{\theta}^{BA} - \text{Id}) = 0$ instead of deviations from $\Phi_{\theta}^{AB} \circ \Phi_{\theta}^{BA} - \text{Id} = 0$.

2 Related Work

Non-parametric Transformation Models and Regularization There are various ways of modeling a transformation between image pairs. The most straightforward non-parametric approach is via a displacement field [35]. Different regularizers for displacement fields have been proposed [20], but they are typically only appropriate for small displacements [24] and cannot easily guarantee diffeomorphic transformations [2]. Fluid models, which parameterize a transformation by velocity fields instead, can capture large deformations and, given a suitably strong regularizer, result in diffeomorphic transformations. Popular fluid models are based on viscous fluid flow [8, 9], the large deformation diffeomorphic metric mapping (LDDMM) model [4] or its shooting variant [23, 37]. Simpler stationary fluid approaches, such as the stationary velocity field (sVF) approach [1, 36] have also been developed. While diffeomorphic transformations are not always desirable they are often preferred due to their invertibility which allows mappings between images to preserve object topologies and prevent foldings which are physically implausible for medical registration tasks. These models have initially been developed for pair-wise image registration where solutions are determined by numerical optimization, but have since, with minimal modifications, been integrated with deep networks [39, 3, 29]. In a deep learning formulation the losses are most frequently the same as for the previous numerical-optimization approaches, but now one no longer directly optimizes over the parameters of the chosen transformation model, but instead over the parameters of a deep network which, once trained, can quickly predict the transformation model parameters.

Fluid registration models are computationally complex as they require solving a fluid equation (either greedily, via direct numerical integration, or via the scaling and squaring approach), but can guarantee diffeomorphic transformations. In contrast, displacement field models are computationally cheaper, but make it more difficult to obtain diffeomorphic transformations. Solution regularity can be obtained for displacement field models by adding appropriate constraints on the Jacobian [15]. Alternatively, invertibility can be encouraged by adding inverse consistency losses, either for numerical optimization approaches [7] or in the context of deep registration networks as is the case for ICON [14]. In particular, ICON showed that inverse consistency alone is sufficient to approximately obtain diffeomorphic transformations when the displacement field is predicted by a deep network. Our work in this article extends ICON by generalizing the inverse consistency loss to a gradient inverse consistency loss, which results in as smooth transformations, faster convergence, and more accurate registration results.

Multi-scale Image Registration Finding good solutions for the optimization problems of image registration is often difficult and one might easily get trapped in an unfavorable local minimum. In particular, this might happen for self-similar images, such as lung vessels, where incorrect vessel alignment might be a locally optimal solution. Further, if there is no overlap between vessels a similarity measure might effectively be blind to misalignment, which is why it is important for a similarity measure to have a sufficient capture range³. Multi-scale approaches have been proposed for optimization-based registration models [33, 22, 34, 40] to overcome these issues. For these approaches, the loss function is typically first optimized at a coarse resolution, and the image warped via the coarse transformation then serves as the input for the optimization at a finer resolution. This approach helps avoid poor local optima as solutions computed at coarser resolutions effectively increase capture range and focus on large-scale matching first rather than getting distracted by fine local details. Multi-scale approaches have also been used for learning-based registration [13, 25, 17, 21, 12, 29] and generally achieve better registration accuracy than methods which only consider one scale [3]. These methods differ in how the multi-scale strategy is incorporated into the network structure and training process. Eppenhof et al. [13] propose a network with multiple sub-networks with each predicting the transformation map given the image pair of the current scale; the final output is a composition of the intermediate maps. To make finding corresponding features that are far away easier, other works [17, 25, 12, 14] warp the input image with the map predicted by the coarse sub-network and then treat it as the new source image of the subsequent sub-network. This increases the capture range of the network itself since it is easier for the sub-networks at the to refine correspondences instead of always starting from the unwarped, albeit downsampled, moving image. However, these methods compute the similarity measure and the regularizer losses at all scales, which requires balancing the weights of the losses across all scales and for each scale between the similarity measure and the regularizer. Hence, there are many parameters

³Note that keypoint approaches [16] and approaches based on optimal transport [30] can overcome some of these issues. However, in this work we focus on the registration of images with grid-based displacement fields.

which are difficult to tune. To side-step the tuning issue, these methods typically use a progressive training strategy to avoid tuning the weights between losses at all scales.

3 Background, Notations, and Inverse Consistency

This section introduces notations and the mathematical background required in the following sections.

Definition 3.1 (Transformation map). A d -dimensional transformation map is a mapping $\Phi : \Omega \rightarrow \Omega$ on a domain $\Omega \subset \mathbb{R}^d$.

Definition 3.2 (Diffeomorphism). A transformation map is a diffeomorphism if Φ and its inverse $\Phi^{-1} : \Omega \rightarrow \Omega$ are both differentiable and a bijection.

Typical optimization-based image registration solves the optimization problem

$$\tau^* = \arg \min_{\tau} \mathcal{L}_{sim}(I^A \circ \varphi_{\tau}^{-1}, I^B) + \lambda \mathcal{L}_{reg}(\tau), \quad (1)$$

where $I^A : \Omega \rightarrow \mathbb{R}$ and $I^B : \Omega \rightarrow \mathbb{R}$ are the moving and fixed images, $\mathcal{L}_{sim}(\cdot, \cdot)$ is the similarity measure, $\mathcal{L}_{reg}(\cdot)$ is a regularizer, and τ are the transformation parameters. In learning-based registration one does not directly optimize over the transformation parameters, τ , but instead over the parameters θ of a neural network (NN) \mathcal{F}_{θ} that predicts $\hat{\tau} \approx \tau^*$ given the moving and fixed images. Such a network is trained over a set of image pairs $I = \{(I_i^A, I_i^B) | i = 1, \dots, N\}$, by solving

$$\theta^* = \arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{sim}(I_i^A \circ (\Phi_{\theta}^{AB})_i, I_i^B) + \lambda \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{reg}((\Phi_{\theta}^{AB})_i), \quad (\Phi_{\theta}^{AB})_i = \mathcal{F}_{\theta}(I_i^A, I_i^B).$$

By training with (I_i^A, I_i^B) and (I_i^B, I_i^A) the loss is symmetric in expectation. For ease of notation, we omit the subscripts i or θ where clear in the following: e.g., we write I^A, I^B and Φ_{θ}^{AB} (or Φ^{AB}).

4 Gradient Inverse Consistency

Picking a regularizer \mathcal{L}_{reg} is important as it implicitly expresses the class of transformation one considers plausible for a registration. Ideally, this plausible transformation space should be learned from the data. However, as non-parametric image registration (at least for image-pairs) is an ill-posed problem, some regularization is beneficial to obtain reasonable solutions. Regularizers frequently involve spatial derivatives of various orders to discourage spatial non-smoothness [20]. Hence, this typically requires picking a type of differential operator (or conversely a smoothing operator) as well as all its associated parameters. Most often this regularizer is *chosen* and not learned from data. Instead of explicitly penalizing spatial non-smoothness the ICON approach advocates using inverse consistency as a regularizer which amounts to learning a transformation space from data in the class of (approximately invertible transforms). To implement inverse consistency, there is a choice of loss. ICON used the sum-of-squares of the composition of the transformations. The ICON regularizer is $\mathcal{L}_{reg}^{ICON} = \|\Phi_{\theta}^{AB} \circ \Phi_{\theta}^{BA} - \text{Id}\|_2^2$, where Id denotes the identity transform, Φ_{θ}^{AB} is the output of the neural network with parameter θ and inputs (I^A, I^B) . GradICON uses the Jacobian of the composition of the maps instead

$$\mathcal{L}_{reg}^{GradICON} = \|\nabla[\Phi_{\theta}^{AB} \circ \Phi_{\theta}^{BA}] - I\|_F^2, \quad (2)$$

where ∇ denotes the Jacobian, I the identity matrix, and the norm $\|\cdot\|_F^2$ denotes the squared Frobenius norm integrated over the domain Ω .

To understand the implicit regularization of the ICON loss, one makes the following modeling choice $\Phi_{\theta}^{AB} = \Phi^{AB} + \varepsilon n^{AB}$ such that $\Phi_{\theta}^{BA}(\Phi_{\theta}^{AB}) = \text{Id}$, i.e. the output of the network is inverse consistent up to a white noise term n with parameter $\varepsilon > 0$ (artificially introduced to make the discussion clear).

In [14], the authors use this white noise to prove that the resulting maps are regularized via a first-order Sobolev norm. Further, they showed in experiments that an approximate diffeomorphism can be obtained without the white noise when used in the context of learning through a deep neural network. Indeed, if the inverse consistency is not exactly enforced, the authors argue, as suggested above, it can be modeled by noise, therefore the resulting smoothness being explained by the theoretical result.

Implicit H^1 regularization of gradient inverse consistency. Since the GradICON loss is on the gradient, it is a natural assumption to put the white noise on the Jacobians themselves rather than on the maps. A direct consequence of this assumption is that $n \ll \nabla n$, i.e. the magnitude of the noise on the gradient is much larger than on the map since it is a white noise integrated in space. In the following, we use $\nabla\Phi_\theta^{AB} = \nabla\Phi^{AB} + \varepsilon\nabla n^{AB}$. We now write the expansion in ε of the gradient inverse consistency loss (2),

$$\mathcal{L}_{reg}^{GradICON} \approx \left\| \left(\nabla\Phi_{\Phi_\theta^{BA}}^{AB} + \varepsilon\nabla n^{AB}(\Phi^{AB} \circ \Phi_\theta^{BA}) \right) \cdot (\nabla\Phi^{BA} + \varepsilon\nabla n^{BA}) - I \right\|_F^2. \quad (3)$$

It remains to expand the term $\nabla\Phi_{\Phi_\theta^{BA}}^{AB} = \nabla\Phi^{AB}(\Phi^{BA}) + \varepsilon\nabla^2\Phi^{AB}(\Phi^{BA})(n^{BA})$. However, this term can be neglected since, as said above, $n \ll \nabla n$. The expansion in ε is only valid at order 1 in ε therefore, inside the L^2 norm, only first order terms need to be taken into account. We have further $\mathcal{L}_{reg}^{GradICON} \approx \|\varepsilon\nabla n^{AB}\nabla\Phi^{BA} + \varepsilon\nabla\Phi_{\Phi_\theta^{BA}}^{BA}\nabla n^{BA}\|_F^2$, since $\Phi^{AB} \circ \Phi_\theta^{BA} \approx I$. Expanding the square leads to

$$\mathcal{L}_{reg}^{GradICON} \approx \varepsilon^2 (\|\nabla n^{AB}\nabla\Phi^{BA}\|_F^2 + \|[\nabla\Phi^{BA}]^{-1}\nabla n^{BA}\|_F^2 + \|\nabla n^{BA}\nabla n^{AB}\|_F^2), \quad (4)$$

since the cross-term involves the product between $\nabla\Phi^{AB}$ and $[\nabla\Phi^{AB}]^{-1}$ which is identity. Finally, we are left with the following loss at first order in ε^2 , $\mathcal{L}_{reg}^{GradICON} \approx \varepsilon^2 \mathbb{E}[\|\nabla n^{AB}\nabla\Phi^{BA}\|_F^2 + \|[\nabla\Phi^{BA}]^{-1}\nabla n^{BA}\|_F^2]$, since the last term only involves the noise term and not the maps it is a constant w.r.t. Φ . This can be further simplified due to the white noise assumption on ∇n^{AB} into

$$\mathcal{L}_{reg}^{GradICON} \approx \varepsilon^2 (\|\nabla\Phi^{BA}\|_F^2 + \|[\nabla\Phi^{BA}]^{-1}\|_F^2). \quad (5)$$

Therefore, this loss is an L^2 regularization of the Jacobian map and the inverse of the Jacobian. **Comparison with ICON.** Interestingly, it is an H^1 type of regularization as in the case of ICON although we could have expected a second-order regularization from the model. We think that this difference is due to the matrix multiplication of the Jacobians in the loss. Although there are several assumptions that can be formulated differently such as the form of the noise and the fact that it is a white noise over the collection of pairs of images (I^A, I^B) , we believe that it is a plausible explanation of the observed regularity of the maps in practice. Importantly, since this regularization is implicit, GradICON as well as ICON can learn based on a slightly more informative prior than this H^1 regularization which relies on simplifying modeling assumptions. Last, there is an important difference with ICON which possibly explains its much faster convergence in practice; since the loss only considers the first derivative of the maps, it is insensitive to the magnitude of the maps contrary to ICON. For instance translations are not seen by this loss, which might help to converge faster. Making the GradICON loss invariant to affine transformations could further improve convergence.

5 GradICON

The gradient inverse consistency loss is general. What remains to be chosen is the network to predict the transformation maps: $\mathcal{F}_\theta(\cdot, \cdot)$. We use a convolutional neural network though other network architectures could, of course, also be used. As discussed in Sec. 2, there are various ways to obtain a multi-scale strategy. GradICON is multi-scale by *composing* multiple transformations at different scales and applying the loss *after* the final composition. Just as for ICON, GradICON only depends on *one* balancing parameter between the similarity measure and the regularizer. However, unlike ICON, GradICON allows using the same balancing parameter across scales.

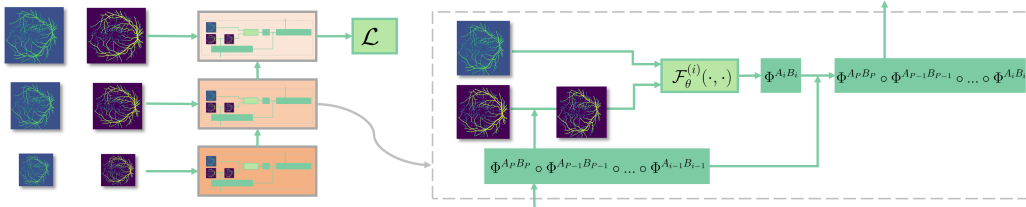


Figure 1: GradICON network structure. The similarity measure is computed only *after* the final transformation prediction making the configuration of P sub-networks independent of the loss. Each sub-network consists of four operations: downsampling, warping, estimating, and composing.

5.1 Network Framework

Fig. 1 shows the GradICON structure, where $\mathcal{F}_\theta(\cdot, \cdot)$ is composed of P sub-networks, denoted $\mathcal{F}_\theta^i(\cdot, \cdot)$, to form a multi-scale approach. The sub-network for $i = 0$ represents the finest resolution. Each sub-network $\mathcal{F}_\theta^i(\cdot, \cdot)$ consists of four operations: 1) **downsampling** the given pair of images by half via average pooling, 2) **warping** the downsampled source image with $\Phi^{A_P B_P} \circ \Phi^{A_{P-1} B_{P-1}} \circ \dots \circ \Phi^{A_i B_i}$, 3) **estimating** $\Phi^{A_i B_i} = u^{A_i B_i} + \text{Id}$ with u being the displacement at the current resolution via a U-Net [28, 10], 4) **composing** $\Phi^{A_i B_i}$ with $\Phi^{A_{i+1} B_{i+1}}$. The final output of $\mathcal{F}_\theta(\cdot, \cdot)$ is

$$\begin{aligned} \mathcal{F}_\theta(I^A, I^B) &= \Phi^{A_P B_P} \circ \Phi^{A_{P-1} B_{P-1}} \circ \dots \circ \Phi^{A_0 B_0}, \\ \Phi^{A_i B_i} &= \mathcal{F}_\theta^i(I^{A_i} \circ \Phi^{A_P B_P} \circ \Phi^{A_{P-1} B_{P-1}} \circ \dots \circ \Phi^{A_i B_i}, I^{B_i}). \end{aligned} \quad (6)$$

Previous multi-scale registration methods use similarity and regularizer losses at each resolution. Instead, GradICON computes a loss only on the map at the final resolution $\Phi^{AB} = \Phi^{A_P B_P} \circ \Phi^{A_{P-1} B_{P-1}} \dots \circ \Phi^{A_0 B_0}$. With this design our loss function is

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{sim}(I^A \circ \Phi^{AB}, I^B) + \mathcal{L}_{sim}(I^B \circ \Phi^{BA}, I^A) + \lambda(\mathcal{L}_{reg}(\Phi^{AB}, \Phi^{BA}) + \mathcal{L}_{reg}(\Phi^{BA}, \Phi^{AB})), \\ \Phi^{AB} &= \mathcal{F}_\theta(I^A, I^B), \Phi^{BA} = \mathcal{F}_\theta(I^B, I^A), \end{aligned}$$

where $\mathcal{L}_{reg} = \mathcal{L}_{reg}^{GradICON}$ defined in Eq. (2) and $\lambda > 0$ is the balancing parameter. \mathcal{L}_{sim} is a similarity measure, e.g., mean squared error (MSE) or normalized cross correlation (NCC) [19].

We note two benefits: 1) the structure of $\mathcal{F}_\theta(\cdot, \cdot)$ is independent of the loss function thus one can adjust both separately and 2) we do not need to balance losses at each scale allowing us to use **only one parameter** in the loss function to balance between the similarity and the regularizer losses.

5.2 Self-Adaptive Multi-Scale Similarity Measure

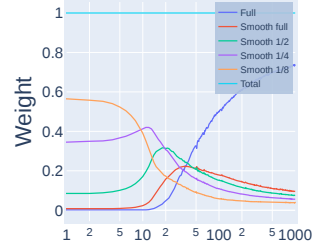
GradICON applies losses only at the highest resolution. In such a design, simple similarity measures (e.g., MSE or NCC) are sufficient for some applications (see our new SOTA for MR knee registration in Sec. 7), but a similarity measure with better capture range is beneficial for more complex structures (e.g., vessels: see Sec. 6.2 for an example). Thus, we propose a **self-adaptive multi-scale similarity measure** which adaptively focuses on different resolutions thereby improving capture range.

Let $\{I^{\tilde{A}(i)} | i = 0, \dots, Q, I^{\tilde{A}(0)} = I^{\tilde{A}}\}$ be the multi-scale pyramid of $I^{\tilde{A}} = I^A \circ \Phi^{AB}$, where $I^{\tilde{A}(Q)}$ is the image at the coarsest resolution. The self-adaptive similarity measure is defined recursively as

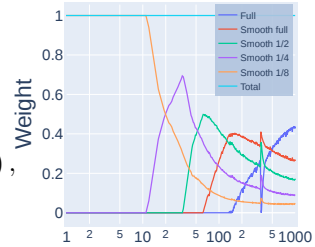
$$\begin{aligned} Sim(I^{\tilde{A}}, I^B) &= g^{(Q)}(I^{\tilde{A}}, I^B), \\ g^{(i)}(I^{\tilde{A}}, I^B) &= \gamma^{(i)} N\tilde{C}C(I^{\tilde{A}(i)}, I^{B(i)}) + (1 - \gamma^{(i)})g^{(i-1)}(I^{\tilde{A}}, I^B), \\ g^{(0)}(I^{\tilde{A}}, I^B) &= N\tilde{C}C(I^{\tilde{A}(0)}, I^{B(0)}), \end{aligned} \quad (7)$$

where $\gamma^{(i)} = N\tilde{C}C(\cdot, \cdot)|_{I^{\tilde{A}(i)}, I^{B(i)}}$ is an adaptive weighting factor through which we do not backpropagate and $N\tilde{C}C = 1 - NCC$. Specifically, $\gamma^{(i)}$ balances between the $N\tilde{C}C$ at the i th scale and $g^{(i-1)}(I^{\tilde{A}}, I^B)$ which is the total weighted similarity measure from the 0th to the $(i-1)$ th scale (i.e., all scales with higher resolution than the i th scale). Hence, the similarity measure focuses on finer scales only if $I^{\tilde{A}(i)}$ and $I^{B(i)}$ are already similar ($\gamma^{(i)}$ gets small).

During the optimization, this self-adaptive multi-scale similarity measure will automatically switch its focus based on the data itself. Fig. 2a shows the behavior



(a) Non-parameter version.



(b) Greedy version.

Figure 2: Weights of the 5 terms $N\tilde{C}C(I^{\tilde{A}(i)}, I^{B(i)})$ with $i = 0..4$ during an optimization-based registration for 1000 iterations. *Full* denotes the weight at full resolution and *total* the sum of the weights.

for the internal weight factors of the self-adaptive similarity measure. For the initial iterations, the weight applied to the similarity measure at the coarsest resolution (smooth $1/8$) dominates the overall similarity measure. Then it progressively switches to finer resolutions.

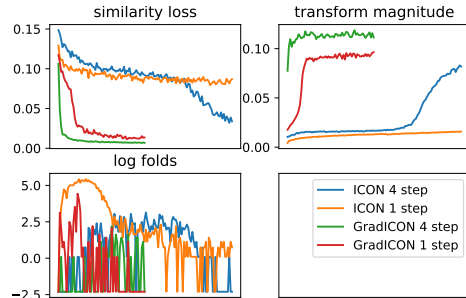
Note that Eq. (7) is parameter-free. When used in some applications, we want the self-adaptive similarity measure to be **greedy**, meaning switching to finer resolution only when $N\tilde{C}C$ at the i th resolution crosses a threshold. This can be controlled by modifying $\gamma^{(i)}$ to

$$\gamma^{(i)} = \text{clamp}\left[\frac{N\tilde{C}C(I^{\tilde{A}(i)}, I^{B(i)})}{h(i; \alpha, \beta)}, 0, 1\right], \quad h(i; \alpha, \beta) = \frac{\alpha}{\beta^{(Q-i)}}, \quad (8)$$

where $h(i; \alpha, \beta)$ is the threshold for the i th scale and is parameterized by α and $\beta > 1$. α is the base threshold and $h(i; \alpha, \beta)$ is reduced as i decreases (moving to finer resolutions).

6 Toy Demo Experiments

6.1 Convergence of Gradient Inverse Consistency



To demonstrate the compounding advantages of our gradient inverse consistency loss and multi-scale network structure, we train 4 networks on a variant of the **Triangles and Circles** dataset [14] with MSE. Specifically, we register outlines of shapes instead of filled shapes. This produces noisier gradients from the similarity measure, drawing out differences between registration methods. We investigate the cross product of (GradICON vs ICON regularization) and (single-step vs multi-step network architecture). Implementation details can be found at Sec. A.1.1.

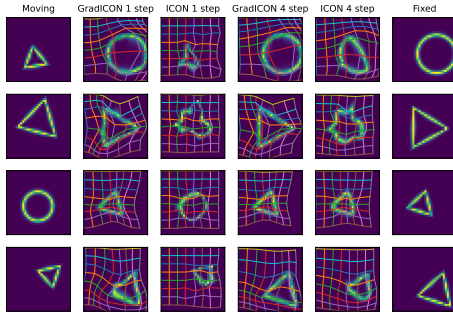


Figure 3: Both multiscale, multistep end-to-end registration and GradICON regularization improve convergence speed; with GradICON regularization having the more profound effect. Top results are plotted over epochs.

our self-adaptive multi-scale NCC (denoted as single, multi_avg and multi_adaptive) as similarity measures. We generate 20 target images with synthetic deformations based on 20 segmentation images from DRIVE. Our 9 combinations are trained for 1,000 epochs using the 20 generated image pairs. Our goal is to study if accurate and well regularised maps can be learned for a tree-like structure, not to assess generalizability. Thus Fig 4 shows results on the **training dataset**.

Observations. Fig. 4a shows that as we increase the number of sub-networks P , the differences between the warped image and the target image reduces, which is quantitatively confirmed by Fig. 4b. One explanation could be that as P increases larger deformations can be captured leading to less error in alignment. This explanation is supported by the quantitative result in Fig. 4c. In addition, compared with other similarity measures, the self-adaptive similarity measure also greatly reduces the number of foldings in the predicted transformation map as shown in Fig. 4d.

Observations This experiment shows that both the GradICON formulation of the inverse consistency loss and the end-to-end multiscale registration approach improve training convergence dramatically on their own and synergistically when both are applied.

6.2 GradICON Ablation Study

This section studies how the network structure and similarity measure affects registrations on the **DRIVE** [32] dataset, which is collection of retinal images and includes blood vessel segmentations. Specifically, we train the GradICON variants based on: 1) number of sub-networks P in the network $\mathcal{F}_\theta(\cdot, \cdot)$, 2) different similarity measures. We set P to 1, 2 and 4 (denoted as net-1, net-2 and net-4) and choose NCC, the average of NCC computed at different scales, and

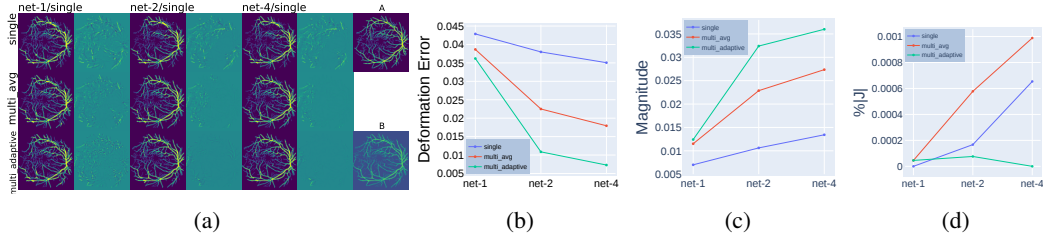


Figure 4: Ablation study of GradICON on the DRIVE dataset. a) Qualitative result for one pair of images from the training set. b) Deformation error of the estimated maps. c) Magnitude of the displacement fields. Units are based on a normalized identity map from $[0,1]$. d) Percentage of pixels with negative determinant of the Jacobian of the estimated maps. Best viewed zoomed.

7 GradICON for OAI registration

We assess registration performance on real 3D magnetic resonance images (MRI) of the knee based on the DESS images of the Osteoarthritis Initiative [26]. A variety of registration methods have been used for the registration of this data many of them requiring affine pre-registration.

Here, we show that GradICON can significantly outperform the current SOTA on the OAI dataset with minimal requirements on parameter tuning and, just as ICON, without the need for affine pre-registration.

Table 1: Evaluation results on OAI dataset. The configurations for the methods are as follows: AVSM(7-Step Affine, 3-Step Deformable), GradICON(2-Step $1/4$ res., 2-Step $1/2$ res., w/o affine). For the high resolution experiment GradICON is configured as GradICON(2-Step $1/4$ res., 2-Step $1/2$ res., 1-Step full res, w/o affine).

Registration method (@ $80 \times 192 \times 192$)	\mathcal{L}_{sim}	DICE	folds
Demons	MSE	63.5	19
SyN	CC	65.7	0
NiftyReg	NMI	59.7	0
NiftyReg	LNCC	67.9	203
vSVF-opt	LNCC	67.4	0
Voxelmorph(w/o affine)	MSE	46.1	83
Voxelmorph	MSE	66.1	39
AVSM	LNCC	68.4	14
ICON	MSE	68.3	118
GradICON	MSE	71.3	2

Registration method (@ $160 \times 384 \times 384$)	\mathcal{L}_{sim}	DICE	folds
NiftyReg	LNCC	70.3	9748
GradICON	MSE	73.3	27

manual segmentations) and 2) the number of folds. The segmentations are not used during training and allow quantifying if the network yields semantically meaningful registrations. Implementation details can be found in Sec. A.2.

Observations. Tab. 1 shows the results. We significantly exceed state-of-the-art. Notably, while we outperform existing methods by registering $1/2$ resolution images as they do, we are also able to register images at the full resolution, achieving further gains. Specifically, this full resolution

Dataset The magnetic resonance (MR) images of the Osteoarthritis Initiative (**OAI**⁴) are used for training and evaluation. Specifically, we use a subset of 2532 images of the full OAI dataset for training and 301 pairs of images for testing. The images are normalized so that the 1th percentile and the 99th percentile of the intensity are mapped to 0 and 1 respectively. Then we clamp the normalized intensity to be in $[0, 1]$. We follow the train-test split used in ICON which can be found in the ICON repository⁵.

Metrics. We evaluate based on the mean **DICE** score between the warped image and the fixed image for the merged segmentations of the femoral and tibial cartilage. We also report the mean number of **folds** in the predicted transformation maps which is computed as the number of voxels where the Jacobian determinant is negative in the transformation map.

Comparing Methods.

We compare GradICON to the methods reported in ICON [14], in terms of 1) cartilage DICE scores between registered image pairs (based on

⁴<https://nda.nih.gov/oai/> w/ Material Transfers Agreement (MTA) regulating permissible data use and consent.

⁵<https://github.com/uncbiag/ICON>

registration was not possible using ICON, because under that regularization strategy, the necessary value of λ exploded as resolution increased; our approach uses the same value of λ for all resolutions.

8 GradICON for COPDGene large deformation lung registration

We demonstrate the ability of GradICON to predict large regular deformations between lung exhale (moving)/inhale (fixed) pairs from the COPDGene dataset within patient. Challenges are: the deformation is complex and large; the visual indication of motion is primarily represented by the deformation of vessels inside the lung which form a complex tree-like structure. This structure makes registrations difficult as the optimization can get easily trapped in a local optimum.

Dataset. We use a subset of 999 inspiratory-expiratory lung CT pairs from the **COPDGene** study⁶ [27] with provided lung segmentation masks for training. CT images are first resampled to isotropic spacing ($2mm$) and cropped/padded evenly along each dimension to obtain a $175 \times 175 \times 175$ volume. Hounsfield units are clamped within $[-1000, 0]$ and scaled linearly to $[0, 1]$. Then the lung segmentation are applied to the images to extract the lung region of interest (ROI). Among the processed data, we use 899 pairs for training and 100 pairs for validation. The **DirLab** dataset [5] is used to evaluate the trained network. It contains 10 pairs of inspiration-expiration lung CTs with 300 anatomical landmarks per pair manually identified by an expert in thoracic imaging. We applied the same preprocessing to DirLab as for COPDGene. The segmentation masks used in preprocessing are computed automatically⁷.

Metrics. The **Mean Target Registration Error (mTRE)** computes the mean distance of corresponding landmarks between a warped moving and a fixed image. We report the mean mTRE over all 10 DirLab pairs in millimeters. Since the landmarks are annotated based on the lung vessels [5], mTRE aims at evaluating the spatial accuracy of the fine structures. The **Percentage of negative Jacobian ($\%|J|$) voxels** is computed as the percentage of voxels with folds in the predicted transformation map. The metrics are evaluated on both transformation maps Φ^{AB} and Φ^{BA} .

Comparing Methods. We compare GradICON with VoxelMorph and LapIRN. VoxelMorph predicts the map via an sVF model and using a single U-Net. LapIRN is the SOTA learning-based method which won the 2020 Learn2Reg challenge [18]. We train LapIRN with two transformation models (displacement field and sVF). Implementation details are in A.3.

Observations Tab. 2 shows that compared with VoxelMorph and LapIRN(sVF) which use more advanced transformation models, GradICON achieves comparable results with LapIRN and outperforms VoxelMorph even though VoxelMorph uses an affine pre-registration. This demonstrates that one can achieve the same level of accuracy with a simpler non-parametric transformation model via gradient inverse consistency and self-adaptive NCC. GradICON outperforms ICON not only in accuracy but also in the regularity of the predicted maps indicating that the gradient inverse consistency loss combined with the self-adaptive NCC is effective.

9 Conclusion

We introduced GradICON, a deep learning registration approach which predicts regular transformation maps for large deformations via gradient inverse consistency. We demonstrated GradICON’s capability to learn accurate registrations on two real registration tasks, MR knee registration and CT lung

Table 2: Evaluation results on DirLab. For each method, registration results are evaluated for two directions: expiration to inspiration (first row) and inspiration to expiration (second row).

Network	\mathcal{L}_{reg}	\mathcal{L}_{sim}	mTRE	DICE	$\% J $
VM (sVF, Aff.)	Diff.	NCC	10.138 9.882	96.90 96.15	0.064 0
LapIRN (sVF)	Diff.	NCC	4.028 2.915	97.16 97.75	0 0
LapIRN (disp)	Diff.	NCC	4.988 4.238	97.00 97.75	0.0062 0.0105
ICON (disp)	—	LNCC	8.917 7.035	96.2 97.2	0.4552 0.3792
GradICON (disp)	—	Adapt.	3.621 3.127	97.58 97.62	0.0044 0.0030

⁶[https://www.ncbi.nlm.nih.gov/gap/w/MTA regulating permissible data use and consent.](https://www.ncbi.nlm.nih.gov/gap/w/MTA%20regulating%20permissible%20data%20use%20and%20consent)

⁷<https://hub.docker.com/r/acilbwh/chestimagingplatform>

registration. In contrast to ICON, GradICON achieves improved convergence speed based on its gradient inverse consistency loss and does not require a scale-dependent scheduling parameter. Hence, GradICON allows for high-quality registrations while relying only on very few tunable parameters, which should make it very attractive for practical use.

Limitations and Future Work. This work explored first order derivative of the inverse consistency loss. It would be interesting to explore how higher order derivatives affect regularity, e.g., by penalizing inverse consistency based on the Hessian. We have also only focused on intensity-based registration methods so far. As discussed in Sec. 2, registration methods based on key-points or optimal transport exist. It would be interesting to combine these approaches with the ideas of GradICON to predict more accurate registration results while retaining spatial regularity.

Potential Negative Societal Impacts. Image registration results might not be accurate or might even fail for certain image pairs in practice. Hence, careful quality control of the registration results should be performed when registrations are used for decision making, e.g., in medical applications. As our experiments are based on knee and lung images only the risk for patient identification is small, but also explicitly prohibited by the MTAs.

Acknowledgments. This work was supported by NIH grants 1R01AR072013, 1R01HL149877, 1R01EB028283, RF1MH126732, R41MH118845 and 5R21LM013670. The work expresses the views of the authors, not of NIH. Roland Kwitt was supported in part by the Austrian Science Fund (FWF): project FWF P31799-N38 and the Land Salzburg (WISS 2025) under project numbers 20102- F1901166-KZP and 20204-WISS/225/197-2019. Data used in this manuscript for the knee registration experiments were obtained from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive. OAI is a collaborative informatics system created by NIMH and NIAMS to provide a worldwide resource for biomarker identification, scientific investigation and OA drug development. Dataset identifier: NIMH Data Archive Collection ID: 2343.

References

- [1] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-Euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer, 2006.
- [2] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.
- [3] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. VoxelMorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.
- [4] M. F. Beg, M. I. Miller, A. Trounev, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.
- [5] R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A. M. Wood, M. S. Ludwig, and T. Guerrero. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Physics in Medicine & Biology*, 58(9):2861, 2013.
- [6] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. F. Frangi. Deep learning in medical image registration. *Progress in Biomedical Engineering*, 3(1):012003, 2021.
- [7] G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE transactions on medical imaging*, 20(7):568–582, 2001.
- [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, 1994.
- [9] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE transactions on image processing*, 5(10):1435–1447, 1996.

- [10] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [11] W. R. Crum, T. Hartkens, and D. Hill. Non-rigid image registration: theory and practice. *The British journal of radiology*, 77(suppl_2):S140–S153, 2004.
- [12] B. D. De Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143, 2019.
- [13] K. A. Eppenhof, M. W. Lafarge, M. Veta, and J. P. Pluim. Progressively trained convolutional neural networks for deformable image registration. *IEEE transactions on medical imaging*, 39(5):1594–1604, 2019.
- [14] H. Greer, R. Kwitt, F.-X. Vialard, and M. Niethammer. ICON: Learning regular maps through inverse consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3396–3405, 2021.
- [15] E. Haber and J. Modersitzki. Image registration with guaranteed displacement regularity. *International journal of computer vision*, 71(3):361–372, 2007.
- [16] L. Hansen and M. P. Heinrich. GraphRegNet: Deep graph regularisation networks on sparse keypoints for dense registration of 3D lung CTs. *IEEE Transactions on Medical Imaging*, 40(9):2246–2257, 2021.
- [17] A. Hering, B. v. Ginneken, and S. Heldmann. mlvirnet: Multilevel variational image registration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 257–265. Springer, 2019.
- [18] A. Hering, L. Hansen, T. C. Mok, A. Chung, H. Siebert, S. Häger, A. Lange, S. Kuckertz, S. Heldmann, W. Shao, et al. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *arXiv preprint arXiv:2112.04489*, 2021.
- [19] G. Hermosillo, C. Chéfd’Hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343, 2002.
- [20] M. Holden. A review of geometric transformations for nonrigid body registration. *IEEE transactions on medical imaging*, 27(1):111–128, 2007.
- [21] Z. Jiang, F.-F. Yin, Y. Ge, and L. Ren. A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration. *Physics in Medicine & Biology*, 65(1):015011, 2020.
- [22] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical image analysis*, 3(4):373–386, 1999.
- [23] M. I. Miller, A. Trounev, and L. Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228, 2006.
- [24] J. Modersitzki. *Numerical methods for image registration*. OUP Oxford, 2003.
- [25] T. C. Mok and A. Chung. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 211–221. Springer, 2020.
- [26] M. Nevitt, D. Felson, and G. Lester. The osteoarthritis initiative. *Protocol for the cohort study*, 1, 2006.
- [27] E. A. Regan, J. E. Hokanson, J. R. Murphy, B. Make, D. A. Lynch, T. H. Beaty, D. Curran-Everett, E. K. Silverman, and J. D. Crapo. Genetic epidemiology of COPD (COPDGene) study design. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 7(1):32–43, 2011.

- [28] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [29] Z. Shen, X. Han, Z. Xu, and M. Niethammer. Networks for joint affine and non-parametric image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4224–4233, 2019.
- [30] Z. Shen, J. Feydy, P. Liu, A. H. Curiale, R. San Jose Estepar, R. San Jose Estepar, and M. Niethammer. Accurate point cloud registration with robust optimal transport. *Advances in Neural Information Processing Systems*, 34:5373–5389, 2021.
- [31] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.
- [32] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4): 501–509, 2004.
- [33] C. Studholme, D. L. Hill, and D. J. Hawkes. Multiresolution voxel similarity measures for MR-PET registration. In *Information processing in medical imaging*, volume 3, pages 287–298. Dordrecht, The Netherlands: Kluwer, 1995.
- [34] P. Thvenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE transactions on image processing*, 9(12):2083–2099, 2000.
- [35] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical image analysis*, 2(3):243–260, 1998.
- [36] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In *International conference on medical image computing and computer-assisted intervention*, pages 754–761. Springer, 2008.
- [37] F.-X. Vialard, L. Risser, D. Rueckert, and C. J. Cotter. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012.
- [38] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim. A survey of medical image registration, 2016.
- [39] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage*, 158:378–396, 2017.
- [40] S. Zhao, Y. Dong, E. I. Chang, Y. Xu, et al. Recursive cascaded networks for unsupervised medical image registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10600–10610, 2019.

A Supplementary Material

A.1 Toy Demo Experiments

A.1.1 Convergence of Gradient Inverse Consistency

Training details Each network is trained using ADAM with a learning rate of 0.001. MSE is used as a similarity metric in all experiments. The GradICON experiments are trained for 64 epochs, while the ICON experiments are trained for 128 epochs to allow them more opportunity to converge. For the GradICON experiments, λ is set to .2, while for the ICON experiments, λ is set to 1024; this setting achieves similar folding rates for both techniques. The 1 step experiments use a single U-Net to predict displacements, while the 4-step experiments use a chain of 4 U-Nets, doubling resolution at each step.

A.2 GradICON for OAI registration

A.2.1 GradICON Design.

We use GradICON based on the composition of up to four U-Nets. We adopt the tallUNet2 in ICON repository as the U-Net implementation. A composition of two U-Nets is initially trained on quarter-resolution image pairs. This network is composed with a third U-Net, trained on half-resolution image pairs. This network is then composed with a fourth U-Net, trained on full-resolution image pairs. We train using ADAM with a batch size of 128 in the quarter-resolution stage, a batch size of 16 in the half-resolution stage, and a batch size of 4 in the full-resolution stage. MSE is used as image similarity measure.

A.2.2 Differences from ICON

This experiment is identical to the ICON training process with these exceptions: We penalize the gradient of the inverse consistency error instead of the error itself. We use a constant value of .2 for λ while ICON ramps λ from 64 up to 6800 during training based on batch foldings. We do not freeze low resolution networks, instead propagating gradients through the whole network at all times. We perform the final step of registration at full resolution for GradICON instead of half resolution, because unlike for ICON this does not produce an explosion in the number of folds.

A.2.3 More results

Fig. 5 shows visual results for two pairs of images from the test set.

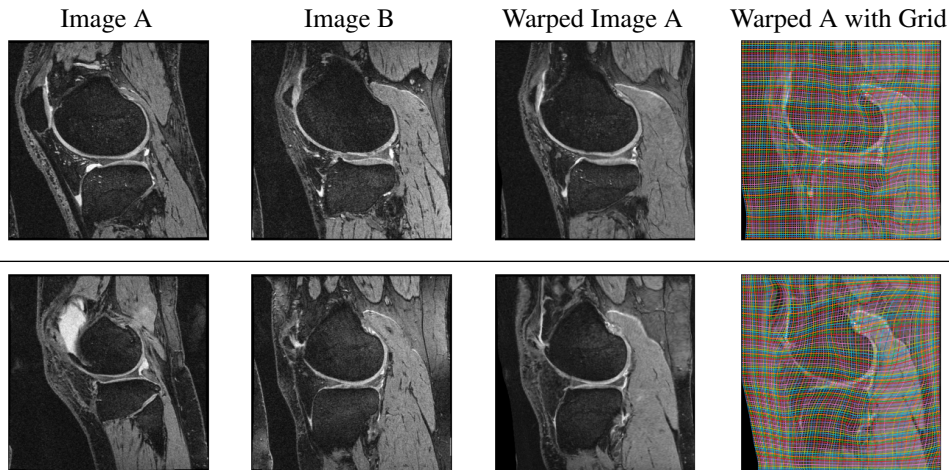


Figure 5: Example registrations performed using GradICON and MSE Similarity on the OAI dataset (test set samples).

A.3 GradICON for COPDGene large deformation lung registration

A.3.1 GradICON Design.

We set $P = 3$ and adopt tallUNet2 in ICON’s official repository as the sub-network. We use our self-adaptive multi-scale similarity measure in this experiment and set $Q = 4$. We train the network for 100 epochs on 4 RTX A6000 GPUs with learning rate set to $1e - 3$ and $\lambda = 0.1$. For the self-adaptive similarity measure, we set $\alpha = 0.1$ and $\beta = 1.3$.

A.3.2 VoxelMorph

We use the official VoxelMorph repository⁸ and train it with the same dataset we used for GradICON. As VoxelMorph requires pre-registration, we train another Affine neural network for pre-registration. Fig. 3 shows the registration accuracy of the Affine network. We use NCC as the similarity measure and set the learning rate to 0.001, and the regularizer weight to 5. We keep the other settings at the default values provided by the official repository. The network is trained on 4 RTX A6000 GPUs in parallel with batch size of 12 for 100 epochs. Each epoch uses the 899 training pairs. Since the official code provides the inverse transformation, the bi-directional registration result is obtained with the forward map and its inverse map computed by the official code.

Table 3: Registration accuracy of the pre-registration Affine network.

	mTRE	Dice
Affine	15.966	80.66
	13.715	80.23

A.3.3 LapIRN

We obtain the network from the official repository⁹ of LapIRN and train the network with the same dataset we used for GradICON. We optimize the hyper-parameters based on the training of our dataset. The network is trained progressively on three different resolutions with the default scheduling provided by the official repository. For each resolution, we need to adjust the learning rate, regularizer factor and window size of the LNCC similarity measure to assure network convergence. Thus, we use the following modifications and hyperparameters. We switched from LNCC to NCC, which helps convergence during training without requiring tuning of the window size for each resolution level. Tab. 4 provides the hyper parameters. We randomly swap the source and target images during training so that the trained network can work for bi-directional registration.

Table 4: Hyperparameters for LapIRN.

Resolution	LapIRN(displacement)		LapIRN(sVF)	
	lr	regularizer	lr	regularizer
1	$1e^{-4}$	0.1	$1e^{-4}$	0.1
1/2	$5e^{-5}$	0.1	$1e^{-4}$	0.1
1/4	$1e^{-5}$	1	$5e^{-5}$	1

A.3.4 More results

This section provides additional evaluation results for the lung registration task. Tab. 5 shows the complete quantitative results including the 25%, 50% and 75% mTRE. In addition, Fig. 6 shows two registration examples from the Dirlab dataset.

⁸<https://github.com/voxelmorph/voxelmorph>

⁹<https://github.com/cwmok/LapIRN>

Table 5: Evaluation results on DirLab. For each method, registration results are evaluated for two directions: expiration to inspiration (first row) and inspiration to expiration (second row).

Network	\mathcal{L}_{reg}	\mathcal{L}_{sim}	mTRE	Percentile			DICE	$\% J $
				25%	50%	75%		
VoxelMorph (sVF, Affine)	Diffusion	NCC	10.138	5.351	9.066	13.818	96.90	0.064
			9.882	6.309	8.930	12.630	96.15	0
LapIRN (sVF)	Diffusion	NCC	4.028	1.584	2.785	5.269	97.16	0
			2.915	1.205	2.043	3.624	97.75	0
LapIRN (disp)	Diffusion	NCC	4.988	2.181	3.940	6.89	97.00	0.0062
			4.238	1.828	3.219	5.799	97.75	0.0105
ICON (disp)	—	LNCC	8.917	5.415	8.050	11.582	96.2	0.4552
			7.035	4.076	6.140	9.235	97.2	0.3792
GradICON (disp)	—	NCC	3.621	1.678	2.790	4.727	97.58	0.0044
			3.127	1.497	2.539	4.102	97.62	0.0030

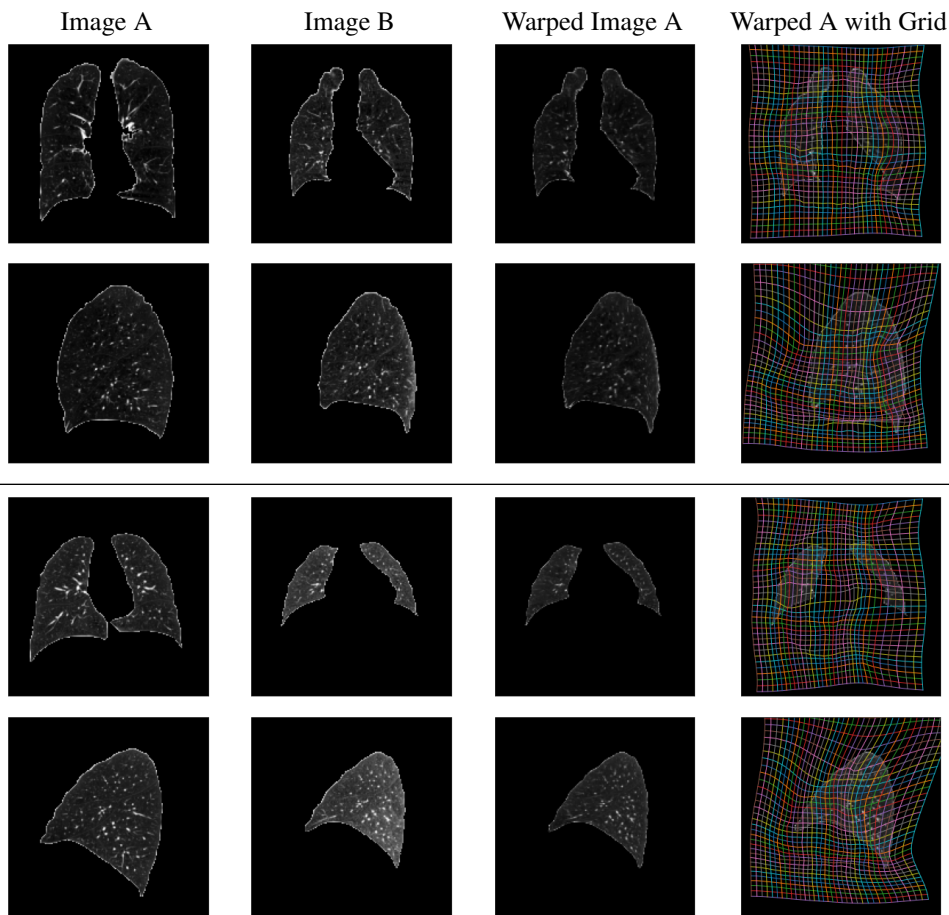


Figure 6: Example registrations performed using GradICON and adaptive multi-scale similarity measure on the DirLab dataset.