Preprints.org

Article

Not peer-reviewed version

Recurrence With Correlation Network for Medical Image Registration

Vignesh Sivan , Teodora Vujovic , Raj Kumar Ranabhat * , Alexander Wong , Stewart Mclachlin , Michael Hardisty *

Posted Date: 30 January 2026

doi: 10.20944/preprints202601.2324.v1

Keywords: medical image registration; neural networks; deep learning

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Recurrence With Correlation Network for Medical Image Registration

Vignesh Sivan [1], Teodora Vujovic [3], Raj Kumar Ranabhat [3], Alexander Wong [1], Stewart Mclachlin [2], Michael Hardisty [3,4,*]

[1] Department of System Design Engineering, University of Waterloo, Waterloo, ON, Canada
[2] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada
[3] Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada
[4] Division of Orthopaedic, Department of Surgery, University of Toronto, Toronto, ON, Canada
* Correspondence: michael.hardisty@sunnybrook.ca

Abstract

This work presents *Recurrence with Correlation Network* (RWCNet), a novel multi-scale recurrent neural network architecture for medical image registration that integrates core principles from optical flow, including correlation volume computation and inference-time instance optimization. In evaluations on the large-displacement National Lung Screening Test (NLST) dataset, which features large displacements, RWCNet exhibited superior performance (total registration error (TRE) of 2.11mm) to deep learning alternatives, and on par results with variational optimization techniques. In contrast, on the OASIS dataset characterized by smaller displacements, RWCNet's results (average dice similarity of 81.7%) were superior to variational optimization techniques and showed a small improvement over other multi-scale deep learning models. Ablation experiments showed that multi-scale features consistently improved performance, where as the correlation volume, number of recurrent steps, and inference-time instance optimization only had large positive impacts on performance in the NLST dataset. The performance of RWCNet compared to approaches that use instance optimzation show that deep learning based methods can find local minima that escape instance optimization methods. The results highlight the need for algorithm hyperparameter selection that adjusts with the dataset characteristics. RWCNet's promising results may imporve registration performance and the speed of computation, allowing many potential applications including, treatment planning, intra-procedural guidance, and longitudinal monitoring.

Keywords: medical image registration; neural networks; deep learning

1. Introduction

Medical image registration describes the task of spatially aligning one medical image to another based on common identifiable features or regions. Medical image registration is commonly required for tasks like image-guided pre-operative planning [1] and quantitative assessments with atlas based segmentation [2], and longitudinal analysis.

Image registration has previously been solved using optimization and variational methods. Deep learning for image registration is an active area of research; deep learning has two potential advantages over existing methods: greatly reduced computational time, and improved accuracy by avoiding local minima. The enhanced speed augments the feasibility of medical image registration for standard clinical applications and offers distinct advantages in time-sensitive settings, including interventional procedures like surgery and focused ultrasound. [3]. Recent investigations have shown improved accuracy of deep learning image registration (DLIR) algorithms over conventional methods [3–5].

DLIR, such as Voxelmorph and Laplacian Image Registration Network (LAPIRN), have shown promise in improving the efficiency and accuracy of image registration by leveraging the power of neural networks. However previous methods do not perform well in all scenarios particularly with

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

large displacements and may not adapt to different datasets' characteristics. Within this study we seek to address this gap by combining three major architectural features: computation of multi-scale features [4], iterative refinement of flow field [6,7] and correlation volume computation [8–10] as well as combine these predictions with inference time optimization [11]. To the authors' knowledge, no prior work has explored a network architecture that combines correlation volume computation, multi-scale refinement and recurrence. This architecture is named *recurrence with correlation* and abbreviated as RWCNet, for its constituent components. The contributions of this work are:

- A novel network architecture for DLIR, *recurrence with correlation network* (RWCNet), that outperforms other continuous-domain and multi-scale networks.
- Investigate importance of architectural components (correlation in feature space, multiple scales, and the recurrent structure) in datasets with large displacements.
- Investigate the effect of inference-time optimization had on performance and found varying results.

2. Background

Conventional approaches to medical image registration formulate registration as an energy minimization problem, where the energy functional quantifies the acceptability of any given solution. If t is a spatial transform function that maps a moving image m to a target (or fixed) image, f, a naive approach to finding an optimal registration would be to find a t that maximizes $\mathcal{S}$, a similarity metric that describes the degree of spatial similarity between the moving and fixed images.

$$t^* = \max_t \mathcal{S}(t(m), f) \tag{1}$$

where t^* is the optimal transform. The moving and fixed images have the same number of dimensions n (i.e., $m, f \in \mathbb{R}^n$). In this section, the number of dimensions is assumed to be 3; thus the term for a position in a 3D image, voxel, is used. For simplicity, it can also be assumed that m and f have the same size.

By itself, the above formulation is ill-posed, since for each of the N voxels in the image, there is an unrestricted number of possible sub-voxel locations ($\gg N$) that the transform can map each of the N voxels to. Thus regularization is needed to restrict the search space for the optimal solution. A common framework is variational optimization, where some regularization functional, $\mathcal{R}$, penalizes unlikely transforms based on prior belief. Smooth and small magnitude displacements are examples of prior beliefs that can be represented and penalized explicitly by regularizers. The objective function in this variational optimization scheme becomes:

$$t^* = \max_t(\mathcal{S}(t(m), f) + \mathcal{R}(t)) \tag{2}$$

Typically, t is a parametric function, belonging to a set of transformations, called the 'transform model', T. In the case of rigid registration, where only translation and rotation are permitted operations, T can be the set of functions parametrized by a homogeneous transform matrix with zero scaling and shearing. When the set of transforms is expanded to include affine transformations, the set of allowable functions might be parametrized as transform matrices with non-zero scaling and squaring. Another commonly used transform model is the displacement field, which describes the displacement of each pixel or voxel in the moving image. For 3D medical images

In the case of deformable registration, the transform can be parameterized by a dense grid, ϕ. The transformation is then a resampling operation of the moving image onto this new grid. The objective function in this scenario becomes

$$\phi^* = \max_\phi(\mathcal{S}(m \circ \phi, f) + \mathcal{R}(\phi)) \tag{3}$$

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

where ϕ is the displacement field, and $m \circ \phi$ denotes the resampling operation. A common way to parameterize the displacement field, ϕ, is to use a dense deformation field, $D \in \mathbb{R}^{3\times H\times W\times D}$, that describes the displacement of each voxel or voxel in a moving image to a fixed image. With this parameterization, the objective function becomes:

$$\phi^* = \max_{\phi}(\mathcal{S}(m \circ (Id + D), f) + \mathcal{R}(D)) \tag{4}$$

where Id is the identity grid and $Id + D$ describes the (sub-) voxel coordinates of each moving voxel in the fixed coordinate frame. One strategy to solve such a problem would be through iterative gradient methods. One can perform simple gradient descent on the loss function described above, using:

$$\phi^t = \phi^{t-1} - \eta \nabla_\phi L \tag{5}$$

where L is the loss function, described by the expression inside the max in Equations (3) and (4), η is the learning rate, and ∇_ϕ is the jacobian operator.

In terms of inputs and outputs, medical image registration is closely related to optical flow, where the objective of the latter is to compute a flow field describing the motion of objects between two images of, typically, the same scene. Optical flow is an important and well-studied computer vision problem and it has a range of applications including action recognition and pose estimation. Common architectural features of optical flow networks include multi-scale features [12] and cost volume layers [9,12,13].

The computation of multi-scale features and cost volumes for 3D medical image registration have been separately explored in prior work [4,10], which have demonstrated improvements in registration performance on standard datasets. This paper introduces a novel, optical flow-inspired network architecture for DLIR, *recurrence with correlation network* (RWCNet). RWCNet combines multi-scale iterative features and a correlation volume computation for medical image registration. To the authors' knowledge, no other work combines these architectural features for DLIR.

3. Related Work

3.1. Voxelmorph

Voxelmorph is one of the earliest and best known methods for DLIR [3]. It trains a 3D U-Net [14] to learn a sub-voxel displacement field, D, that jointly optimizes image fidelity of the transformed image, denoted $\mathcal{S}$, and a regularization loss promoting smooth spatial transformation, denoted $\mathcal{R}$. The loss functional, L, describing the goodness of fit for registering a moving image m to a fixed image f is expressed as:

$$L = \mathcal{S}(m \circ (Id + D), f) + \mathcal{R}(D) \tag{6}$$

The resampling operation is carried out differentiably using the spatial transformer network (STN) [15]. The loss function described by (6) can be augmented with the correspondence of auxiliary data such as segmentations and keypoints. Voxelmorph was found to be competitive with optimization-based registration methods such as Demons [16] and Symmetric Normalization [17].

3.2. Laplacian Image Registration Network

Laplacian Pyramid Image Registration Network (LapIRN) [4] learns flow fields at $N = 3$ resolutions. At each resolution a CNN with residual skip connections learns a displacement field by jointly optimizing an image fidelity term and smoothness term described by (6). LapIRN is trained in a coarse to fine manner; thus the CNN's at low resolutions are trained before training the networks responsible for learning at higher resolutions. The flow fields learned at low resolutions are upsampled and used to warp the moving input image at higher resolutions.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

Multi-scale refinement operates on the principle that an optimal displacement field at low resolution is also a good displacement field at high resolution [4]. Moreover, at coarse resolution the optimization problem is simpler, owing to the fact that the required displacements are a smaller number of the coarse voxels and there are fewer high-level features to match; refining the flow field from a coarse-to-fine resolution can thus simplify the optimization problem. LAPIRN showed improvement in the large-displacement setting, when compared to Voxelmorph and conventional approaches. LapIRN uses convolutional neural network (CNN) architecture with residual connections to learn the flow fields at each resolution.

The method presented in this paper also learns flow fields from coarse to fine resolutions, like LapIRN. RWCNet, the architecture presented in this paper, differs from the one used by LapIRN; a recurrent CNN architecture with features encoders and a correlation volume layer is used.

3.3. ConvexAdam

ConvexAdam [11] combines non-parametric discrete optimization with continuous-domain instance optimization and is able to achieve achieve state-of-the-art performance in several large displacement datasets. RWCNet shares two features with ConvexAdam: they both compute MIND (Modality Invariant Neighborhood Descriptor) features [18] of the fixed and moving image. Second they both use continuous domain optimization as shown in (5) for the final step.

The ConvexAdam algorithm uses an initial discrete optimization whereas RWCNet uses recurrent neural networks to determine the displacement field (discussed below).

3.4. Recurrent Neural Networks for Registration

Recurrent neural networks for image registration have been studied in the 2D [7] and 3D [6] registration settings. In Sandkuhler et al., an RNN composed of a Gated Recurrent Units (GRU) [19] was used to estimate parameters for a parametric Gaussian transformation function. At each time step, the transformation was applied to the moving image, in order to generate a new moving image which was fed into the new time step.

3.5. Recurrent All-Pairs Field Transforms for Optical Flow (RAFT)

Optical flow differs from the general image registration problem because the images are typically of the same object and the consistency of the intensity of objects between both objects is expected. The general image registration problem does not have these constraints. Nevertheless, the two problems are similar. The inputs to the algorithms are two images the output is a flow field from one image to the other. As such, optical flow algorithms can be used directly for medical image registration.

The architecture of RWCNet is inspired by RAFT [13]. The RAFT network architecture has three key characteristics: 1) CNN feature encoders, 2) correlation volume computation at multiple scales between all pairs of pixels in the feature tensor and 3) recurrent CNN architecture for computing and refining the flow field and performing a 'lookup' that subsamples the global cost volume. The computation of a correlation volume from feature encoders is a common architectural structure of many optical flow networks [8,9,12] and is based on the notion that the correlation volumes force the network to learn salient features for both input images, because the same object is expected to appear in both images. This might not be happening in other formulations when the inputs are stacked together. The correlation volume (C) can be computed from the feature vectors (F_f and F_m) as:

$$C(i,j,k,p,q,r) = \frac{F_f(i,j,k) \cdot F_m(p,q,r)}{\sqrt{n_{\text{features}}}} \tag{7}$$

RWCNet differs from RAFT in several ways. RWCNet employs strategies for decreasing the computational complexity. RAFT computes a cost between all pairs of voxels, which is prohibitively expensive for 3D volumes, as such RWCNet instead considers correlation volumes for whole down sampled images, and patches of the image at higher resolutions. Similiarly computing the correlation volume at multiple resolutions simultaneously as is done in RAFT is also computationally expensive.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

As such, RWCNet considers each resolution in isolation, similiar to the strategy used in LapIRN for DLIR or PWCNet[12] for optical flow.

4. Methods

4.1. Coarse to Fine Registration.

We adopt a coarse to fine approach to image registration. Figure 1 provides an overview of this approach. For each resolution, s, a new RNN is trained using inputs from the previous resolution. The weights from previous resolutions are frozen at finer resolutions. At fine resolutions computing the correlation volume for the whole volumes becomes prohibitively expensive; as such, our approach divides the input images into uniform, non-overlapping windows or patches. The size of the patches at each resolution is parameterized by the 'patch factor', $p^s \in [0, 1]$. The size of the patches at resolution s is computed as $p^s \times S$ where S is the size of the full image at $1\times$ resolution. At higher resolutions, the flow field is used to warp the initial moving image (or patch) using flow fields computed at lower resolutions. Furthermore, the final hidden state is cached at lower resolutions and added to the initial hidden state at higher resolutions, increasing the non-linearity of the network and providing additional context to the network.

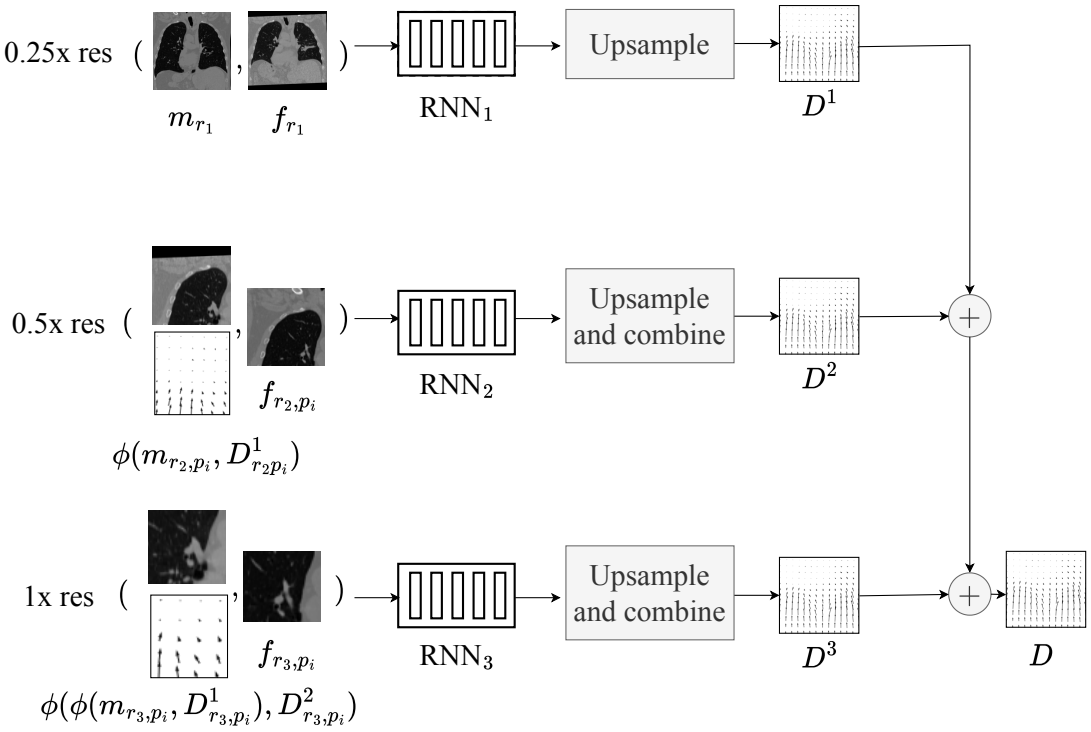


Figure 1. Multi-scale refinement of flow field using separate RNNs for each resolution. Note, at higher resolutions, patches of the input image are fed into the network, rather than the entire image. Not shown is the fact the final hidden layer is upsampled and concatenated with the first hidden state of the next resolution.

4.2. Sub-network Architecture.

The architecture for the recurrent CNN sub-network is shown in Figure 2. Given a fixed and moving image pair (f and m, respectively), the network first learns fixed and moving features (F_f and $F_{m,0}$) by feeding both images through a feature extractor network. A voxel-wise correlation between the fixed and moving features is computed, C_0. Due to the large number of dimensions, the correlation is restricted so that only voxels within a certain range, r of the moving voxel are considered. Additionally, the moving image is fed through a context network that extracts contextual information for the hidden network. The output of the contextual network is used as the initial hidden state of the RNN, h_0. Finally, a displacement field with zero displacement, D_0 is initialized.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

The hidden state, flow, correlation volume and moving image are fed into an update block that is a modified gated recurrent unit (GRU) [19]. The GRU is similar to the one implemented in RAFT [13] but in 3D, and it outputs a new hidden state and a new displacement field, h_1 and ΔD. The new displacement field is used to update the aggregate displacement, i.e, $D_1 = \Delta D + D_0$. This new displacement field is used to warp the moving features (generating $F_{m,1}$), which can be used to generate a new correlation volume, C_1, for the next RNN time step. This process of updating displacement field with GRU cell and generating the correlation volume is repeated for N RNN layers.

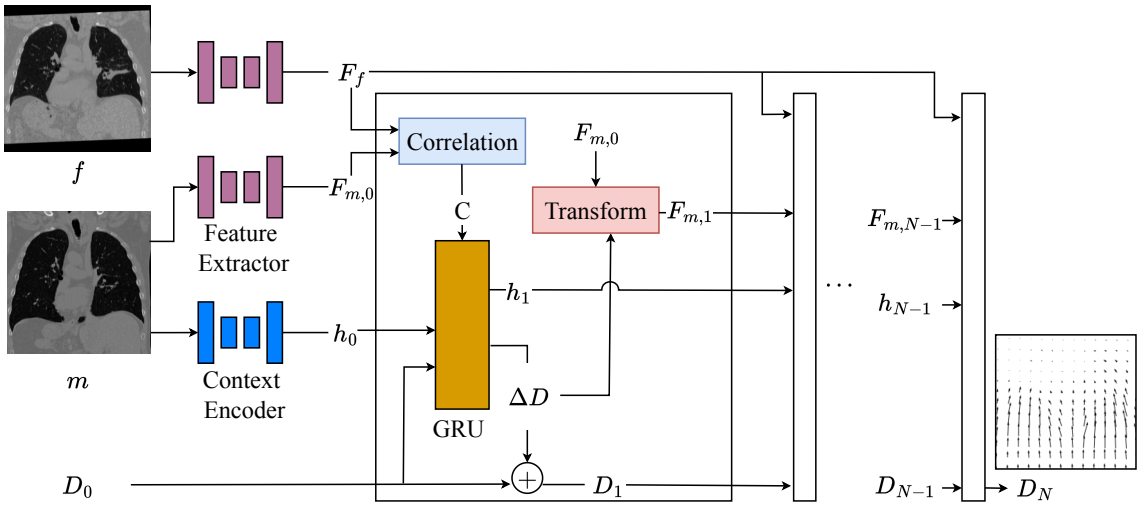


Figure 2. RNN Sub-network architecture. It has 3 main components: 1) feature extractor and context encoder networks for extracting fixed and moving features as well as a context 2) computation of a cost volume through correlation of the input features 3) update aggregate displacement using a GRU-based update block.

4.3. Ablation Experiments

Ablation experiments were performed to assess the contribution of the individual architectural features. To test the impact of multi-scale refinement, the network was trained to learn registration fields at a single resolution. To study the impact of correlation, the correlation volume computation was removed; instead the input to the GRU was simply a concatenation of the input features, which was consistent with Voxelmorph and LapIRN.

5. Experiments

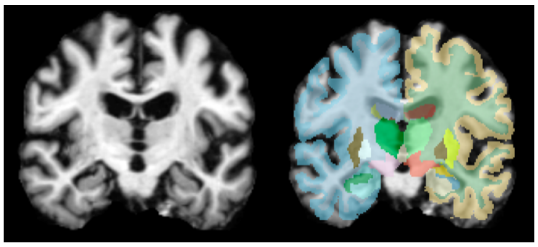
5.1. Datasets

Experiments were performed with the OASIS [20] and NLST [21] datasets prepared for the MICCAI 2022 Learn2Reg workshop challenge [22] (Figure 3). The OASIS dataset consists of 414 T1-weighted MRI scans of individuals from ages 18-96 with mild to severe Alzheimer's. The scans are skull-stripped and resampled onto an isotropic grid and cropped to a uniform size. 35 segmentation labels are provided for important brain regions. The dataset is split into 395 images for training and 19 for validation. Intersubject registration in this context could be used for constructing a sub-population brain atlas or for analysing intensity changes in consistent brain regions that are linked to disease progression.

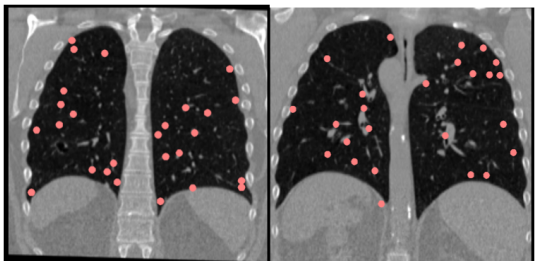
NLST is a lung-CT dataset with pairs of inhale/exhale scans; keypoints and masks are provided by the Learn2Reg challenge for semi-supervised training. We used a subset of the image pairs (100 out of 150) of the NLST dataset released by the Learn2Reg challenge for training and validation, with a 90:10 training/validation split. Since respiration is accompanied by a large change in lung volume, the displacement field required to register NLST is large, relative to OASIS.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1



(a) Oasis



(b) NLST

Figure 3. Sample images from the OASIS (a) and NLST (b) datasets. Note, segmentations are available for the OASIS dataset and keypoints indicating the positions of nodules are available for the NLST dataset.

5.2. Training Parameters

The subnetworks at each scale were trained separately from coarse to fine. At higher resolutions, corresponding patches of the fixed and moving images are passed into the network to decrease GPU memory requirements. Table 1 shows the number of steps and the sizes of the inputs at each resolution. To address overfitting, dropout with probability 0.5 was used for the feature networks. The network took about 30 hours to train both on the NLST and OASIS datasets on an NVIDIA A100 with 40GB of RAM.

Distinct composite loss functions were used for the OASIS $L_{\text{OASIS}}$and NLST L_{NLST} datasets. An MSE term was used to capture difference between the warped moving image I_{moving} and the fixed image I_{fixed} intensities:

$$\text{MSE} = \frac{1}{N} \sum_{v \in \text{voxels}} (I_{\text{fixed},v} - I_{\text{moving},v})^2 \tag{8}$$

The Dice loss term was used to measures the similarity between the warped segmentation S_{moving} and the fixed segmentation S_{fixed}:

$$L_{\text{DICE}} = 1 - \frac{2 \sum_{v \in \text{voxels}} S_{\text{fixed},v} S_{\text{moving},v}}{\sum_{v \in \text{voxels}} S_{\text{fixed},v} + \sum_{v \in \text{voxels}} S_{\text{moving},v}} \tag{9}$$

A regularization loss term used that average magnitude of the gradient of the flow field:

$$\mathcal{R} = \frac{1}{N} \sum_{v \in \text{voxels}} |\nabla \mathbf{D}_v| \tag{10}$$

The Total Registration Error (TRE) measures the discrepancy in millimeters between corresponding keypoints in the fixed and moving images. Given m pairs of corresponding keypoints $\mathbf{p}_{\text{fixed},i}$ and $\mathbf{p}_{\text{moving},i}$ in the fixed and moving images, respectively, the TRE is computed as

$$\text{TRE} = \frac{1}{m} \sum_{i=1}^{m} \left\| \mathbf{p}_{\text{fixed},i} - \mathbf{p}_{\text{moving},i} \right\| \tag{11}$$

For OASIS, MSE was combined with Dice loss and the regularization in a weighted (w_{MSE}, w_{TRE}, and w_R) sum:

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

Table 1. Resolution-specific Parameters.

Resolution	RNN Steps	Patches Per Image	Training Steps
0.25	12	1	30000
0.5	12	8	45000
1	4	8	60000

$$L_{\text{OASIS}} = w_{\text{MSE}} \cdot \text{MSE} + w_{\text{DICE}} \cdot L_{\text{DICE}} + w_{\mathcal{R}} \cdot \mathcal{R} \tag{12}$$

For NLST, the data was range normalized to between 0 and 1, with -4000 serving as the minimum value and 16000 serving as the maximum value. The NLST loss function was a weighted (w_{MSE}, w_{TRE}, and w_R) summation of the MSE, TRE and regularization term based on the mean gradient of the flow field:

$$L_{\text{NLST}} = w_{\text{MSE}} \cdot \text{MSE} + w_{\text{TRE}} \cdot \text{TRE} + w_{\mathcal{R}} \cdot \mathcal{R} \tag{13}$$

For both datasets, an Adam optimizer with learning rate of 3×10^{-4} was used.

LAPIRN, another multi-resolution model, was trained on OASIS using training parameters from [4]. We use the non-diffeomorphic variant, which does not ensure topological consistency of the spatial transformation, but achieves greater quantitative accuracy. For NLST, we augment training with a supervised discrepancy loss between the fixed and moving keypoints once the displacement is applied to the moving keypoints. We use a MSE loss instead of the normalized cross correlation (NCC) loss prescribed by the original paper to be consistent with the the loss used for training RWCNet.

6. Results

Figure 4 shows qualitative results generated for the NLST and OASIS datasets. Table 2 summarizes the results when comparing RWCNet with LAPIRN on the NLST and OASIS datasets. RWCNet outperforms LAPIRN on both datasets. However, the difference in Dice is quite modest 0.7%. For the NLST dataset, the difference in performance is much more pronounced, with the difference in average TRE being >3mm. These results suggest that the architectural features of RWCNet, significantly aid generalization performance in the more challenging large-displacement setting.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

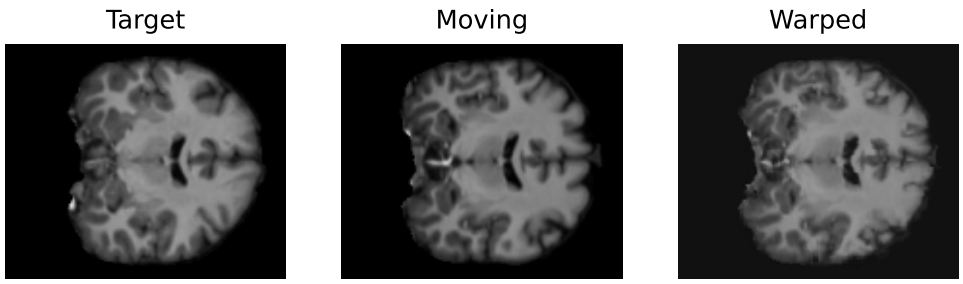
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

Table 2. Experiment Results on NLST and OASIS Validation

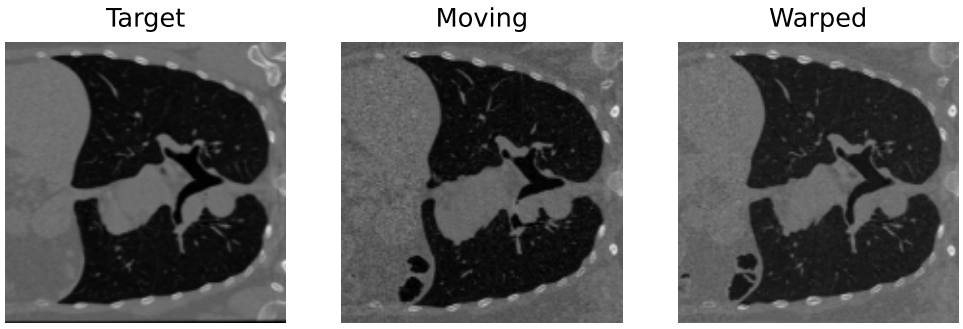
Experiment	NLST TRE (mm) ↓	OASIS Dice (%) ↑
Zero Displacement	9.73	52.4
ConvexAdam [11]	1.48	76.4
ConvexAdam w/o IO [11]	2.78	75.6
LAPIRN [4]	5.51	80.0
RWCNet	**2.11**	**80.7**
RWCNet with Adam IO	**1.48**	**76.4**

Table 3. Ablation Experiment Results on NLST and OASIS Datasets

Experiment	NLST TRE (mm) ↓	OASIS Dice (%) ↑
Baseline	2.11	80.7
At single resolution (4x)	5.52	74
Without correlation	4.10	80.0
2 RNN timesteps	5.17	80.1



(a) OASIS qualitative results



(b) NLST qualitative results

Figure 4. Qualitative results, showing the target, moving and warped images for OASIS (a) and NLST (b).

These ablation results (Table:3) provide interesting insights into the role that architecture plays in registration accuracy in different datasets. Unsurprisingly, multi-resolution registration plays a crucial role in the accuracy of RWCNet; registering at only 4× downsampling on the NLST dataset yields a keypoint discrepancy of 5.52 mm, whereas registering at multiple resolutions yields a discrepancy of 2.11mm. OASIS Dice, likewise, drops from 80.7% to 74.0%.

The impact of correlation and number of RNN time steps was markedly different for both datasets. In the OASIS datasets, replacing correlation with stacking of the input feature tensors did not drastically

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

impact the registration performance. The Dice score dropped by 0.7. Likewise, when only 2 RNN time steps were used in RWCNet, the drop in accuracy was even lower in the OASIS dataset; the Dice score only dropped by 0.6%. This was in contrast to NLST, where decreasing the number of RNN time steps and removing correlation drastically decreased performance. The keypoint accuracy decreased to 4.10mm when correlation was not computed. Likewise, when only 2 time steps were used, the accuracy decreased to 5.17mm. Instance optimization, which fine tuned the flow field using an ADAM optimizer, did improve the performance of RWCNet in NLST dataset so that it was on par with ConvexAdam (that also uses instance optimization). Interestingly the performance of RWCNet was degraded in the OASIS dataset by instance optimization, again to a performance that was on par with ConvexAdam.

The intrinsic dataset attributes(inter vs intra subject or large vs small displacements), might have contributed to the observed variations in performance within the ablation study. OASIS has relatively small displacements compared to NLST and as such may not benefit as much from correlation volumes and the RNN structure. In contrast both datasets showed that a multi-resolution approach improved performance. These observations could be helpful for designing generalizable methods for image registration problems.

7. Conclusions

The DLIR network developed in this investigation, RWCNet, borrows key ideas from optical flow literature; specifically that of correlation volume computation and iterative refinement, to improve medical image registration performance. For the large displacement NLST dataset, results with instance optimization were the same with either RWCNet or ConvexAdam implying that the steps other than instance optimization did not influence the final output for those setups. Without instance optimization it was found that RWCNet was more performant than comparable deep learning methods. In the small displacement setting, RWCNet was found to have a very small improvement over LAPIRN another deep learning approach and improved on improved over variational approaches. More tuning of algorithm parameters is need if instance optimization is to be incorporated into registration algorithms with deep learning. Future work should investigate whether the instance optimization step can be augmented with features learned by the network to ensure a more uniform improvement in all datasets.

Ablations illustrated the benefits of the adopted methods. Multi-scale iterative refinement played a key role; performing registration at a single resolution drastically decreases the accuracy of the model. RNN steps and the computation of the explicit correlation volume were notably influential, their primary effect was observed in the dataset with large displacements. On OASIS, the Dice score did not change drastically without these features. This suggests that these architectural features were valuable in large displacement settings. This was a significant finding, as it can inform future work for large displacement registration.

The findings in this paper highlight the need for dataset-specific hyperparameter selection. In the small displacement setting RWCNet yields only a modest improvement over LapIRN. In the larger displacement setting, RWCNet drastically outperformed LapIRN. RWCNet was more resource intensive than LapIRN therefore there was little benefit to using RWCNet for cases with small displacements. The findings of this work were consistent with the concept of automatic self-configuring networks that vary their architecture components based on the dataset. Such an approach for image registration would be similar to nnU-net [23] for medical image segmentation, this would be in contrast to variational image registration that have been distributed as one size fits all with parameters adjustable by the user. The approach has many potential application where fast computation is needed such as intra-procedural registration, further there are applications where variational approaches do not converge to the global optimal registration trasform.

Author Contributions: Our team includes 6 researchers that authored the manuscript. Michael Hardisty made significant intellectual contributions specifically to the theoretical development, experimental design, the inter-

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

pretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article. Stewart McLachlin made significant intellectual contributions specifically to the theoretical development, experimental design, the interpretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article. Alexander Wong made significant intellectual contributions specifically to the theoretical development, experimental design, the interpretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article. Teodora Vujovic made significant intellectual contributions specifically to the prototype development, the interpretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article. Raj Kumar Ranabhat made significant intellectual contributions specifically to the prototype development, the interpretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article. Vignesh made significant intellectual contributions specifically to the prototype development, the interpretation and analysis of the data, they contributed to the writing of the article and have approved the final version of the article.

Funding: This research was funded by INOVAIT through the Government of Canada's Strategic Innovation Fund, Feldberg Chair for Spinal Research, and Biotalent Canada.

Data Availability Statement: The code for RWCNet is available at https://github.com/vigsivan/RWCNet.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RWCNet	Recurrence with Correlation Network
NLST	National Lung Screening Test
TRE	Total Registration Error
DLIR	Deep Learning Image Registration
CNN	Convolutional Neural Network
LAPIRN	Laplacian Image Registration Network
MIND	Modality Invariant Neighborhood Descriptor
RNN	Recurrent Neural Networks
GRU	Gated Recurrent Unit

References

1. Risholm, P.; Golby, A.J.; Wells, W.M. Multi-Modal Image Registration for Pre-Operative planning and Image Guided Neurosurgical Procedures. *Neurosurgery clinics of North America* **2011**, *22*, 197–206. [CrossRef].
2. Hardisty, M.; Gordon, L.; Agarwal, P.; Skrinskas, T.; Whyne, C. Quantitative characterization of metastatic disease in the spine. Part I. Semiautomated segmentation using atlas-based deformable registration and the level set method. *Medical physics* **2007**, *34*, 3127–3134.
3. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging* **2019**, *38*, 1788–1800. [CrossRef].
4. Mok, T.C.W.; Chung, A.C.S. Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks, 2020. arXiv:2006.16148 [cs, eess].
5. Heinrich, M.P.; Hansen, L. Voxelmorph++ Going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation, 2022. arXiv:2203.00046 [cs].
6. Zhao, S.; Dong, Y.; Chang, E.I.C.; Xu, Y. Recursive Cascaded Networks for Unsupervised Medical Image Registration. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 10599–10609. [CrossRef].
7. Sandkühler, R.; Andermatt, S.; Bauman, G.; Nyilas, S.; Jud, C.; Cattin, P.C. Recurrent Registration Neural Networks for Deformable Image Registration, 2019. arXiv:1906.09988 [cs, stat].
8. Fischer, P.; Dosovitskiy, A.; Ilg, E.; Häusser, P.; Hazırbaş, C.; Golkov, V.; van der Smagt, P.; Cremers, D.; Brox, T. FlowNet: Learning Optical Flow with Convolutional Networks, 2015. [CrossRef].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2026 doi:10.20944/preprints202601.2324.v1

9. Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks, 2016. arXiv:1612.01925 [cs].
10. Heinrich, M.P. Closing the Gap between Deep and Conventional Image Registration using Probabilistic Dense Displacement Networks, 2019. arXiv:1907.10931 [cs].
11. Siebert, H.; Hansen, L.; Heinrich, M.P. Fast 3D Registration with Accurate Optimisation and Little Learning for Learn2Reg 2021. In *Biomedical Image Registration, Domain Generalisation and Out-of-Distribution Analysis*; Aubreville, M.; Zimmerer, D.; Heinrich, M., Eds.; Springer International Publishing: Cham, 2022; Vol. 13166, pp. 174–179. [CrossRef].
12. Sun, D.; Yang, X.; Liu, M.Y.; Kautz, J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume, 2018. [CrossRef].
13. Teed, Z.; Deng, J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow, 2020. [CrossRef].
14. Çiçek, C.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, 2016. [CrossRef].
15. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial Transformer Networks, 2016. arXiv:1506.02025 [cs].
16. Thirion, J.P. Non-rigid matching using demons. In Proceedings of the Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, pp. 245–251. [CrossRef].
17. 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. *Medical Image Analysis* **2008**, *12*, 26–41. [CrossRef].
18. Heinrich, M.P.; Jenkinson, M.; Bhushan, M.; Matin, T.; Gleeson, F.V.; Brady, S.M.; Schnabel, J.A. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis* **2012**, *16*, 1423–1435. [CrossRef].
19. Cho, K.; van Merrienboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, 2014. [CrossRef].
20. 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* **2007**, *19*, 1498–1507. [CrossRef].
21. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* **2011**, *258*, 243–253. [CrossRef] [PubMed].
22. Hering, A.; Hansen, L.; Mok, T.C.W.; Chung, A.C.S.; 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. *IEEE Transactions on Medical Imaging* **2023**, *42*, 697–712. [CrossRef] [PubMed].
23. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation | Nature Methods.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.