Continuous Optimization for Medical Image Registration of Large Displacement Datasets

Teodora Vujovic — Sunnybrook Research Institute
Vignesh Sivan — University of Waterloo
Michael Hardisty — Sunnybrook Research Institute
Stewart McLachlin — University of Waterloo
Email: tvujovic@uwaterloo.ca

Abstract

Medical image registration is an important component of many clinical analysis pipelines. While this approach has conventionally been approached using optimization, deep learning has recently gained interest due to speed of inference. In this work, we demonstrate that gradient-based optimization with modified learning rate is sufficient to achieve state of the art accuracy in a large displacement dataset. The approach is also competitive in terms of speed with deep learning approaches. We discuss the implications of this work on medical image registration at large and future directions to extend the findings presented here.

1 Introduction

Medical image registration (MIR) is the task of aligning a pair of medical images. It has a myriad of clinical applications including surgical planning, quantitative disease tracking and medical image atlas creation. Conventional methods for MIR typically formulate registration as an optimization problem. While conventional methods have been successful, they are typically slow. There has recently been interest in applying deep learning to the image registration problem, which can produce registration orders of magnitude faster than conventional methods.

While deep learning methods have demonstrated the ability to produce accurate registrations on a range of datasets, current approaches suffer from several limitations. Firstly, deep learning requires large datasets to train a model. In the space of medical image analysis, obtaining such large datasets is often difficult. Secondly, recent deep learning methods have been shown to not perform as well as conventional registration methods where there is a large spatial disparity between the images that are being registered.

To address the large displacement issue, recent methods have adopted a hybrid approach to registration, combining both deep learning with conventional opimization for finetuning. ConvexAdam, a more recent approach yet, has shown that combining discrete optimization with continuous optimization as finetuning can achieve state of the art performance on large displacement dataset. The discrete optimization, it is postulated, enables the optimizer to overcome local optima.

In this work, we show that continuous optimization, with the correct choice of hyper-parameters, is able to achieve state of the art performance on a large displacement lung CT dataset, outperforming state of the art and many contemporary deep learning approaches. We show that continuous optimization on the GPU is also fast, overcoming the performance limitations of past optimization approaches.

2 Background

The objective function for finding the optimal displacement field, ϕ^*, for registering a moving image, m, to a fixed image, f, typically has the form:

$$\phi^* = \max_{\phi}(\mathcal{S}(m \circ \phi, f) + \mathcal{R}(\phi)) \tag{1}$$

where $\mathcal{S}$ is some measure of similarity between the transformed moving image and the fixed image and $\mathcal{R}$ is some regularization functional that penalizes highly irregular or complex transformations. Typical similarity measures include mutual information, normalized cross correlation and MIND, described by [1]. A common regularizer is the $L2$-norm of the displacement field or that of its gradient.

A common way to parameterize the displacement field, ϕ, is to use a dense deformation field, $D \in R^{3\times H\times W\times D}$,that describes the displacement of each pixel or voxel in a moving image to a fixed image. With this parametrization, Equation (1) becomes

$$\phi^* = \max_{\phi}(\mathcal{S}(m \circ (Id + D), f) + \mathcal{R}(D)) \tag{2}$$

where Id is the identity grid and $Id + D$ describes the (sub-) voxel coordinates of each moving voxel in the fixed coordinate frame. The transform function, $m \circ \phi$ then, is simply a resampling operation that resamples and interpolates the moving image to a new coordinate grid. For deep learning, this resampling operation can be carried out using spatial transformer networks, which formulate the resampling as a differentiable operation, enabling backpropagation of error.

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:

$$\phi^t = \phi^{t-1} - \eta \nabla_{\phi} L \tag{3}$$

where L is the loss function, described by the expression inside the $\max$ in Equations (1) and (2), η is the learning rate, and ∇_{ϕ} is the jacobian operator. In general, the loss objective is non-convex and simple gradient descent will converge to a local optimum. Moreover, simple gradient descent tends to oscillate as it always follows the steepest slope. A common strategy, especially when optimizing neural networks, is to dampen the learning rate with some measure of momentum such that the weights do not move in directions where there is large oscillation, hence low momentum. Adam [2] is a common optimizer that uses momentum to dampen oscillations. Another common strategy is to use a learning rate scheduler to decrease the magnitude of η as the optimizer begins to converge, so that there is oscillation closer to the optimal point.

2.1 ConvexAdam

Naive gradient descent using Equation (3) can converge to local optima. ConvexAdam combines two strategies for overcoming local optima: dual optimiztation and discrete search space. Shown in Equation (4) is the formulation that ConvexAdam uses. ConvexAdam [3] makes use of the principal of duality in convex optimization by introducing a dual variable **v** and reformulating the loss function as:

$$L'(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}) + \frac{1}{2\theta}(\mathbf{v} - \mathbf{u})^2 + R(\mathbf{u}) \tag{4}$$

This optimization can be solved in a coupled fashion; in each iteration **u** is first optimized, followed by **v** and the parameter θ that enforces the consistency between **u** and **v** is decreased according to some schedule. When decoupled, the optimization of **u** and **v** becomes convex and can be solved using convex optimization techniques.

Further, ConvexAdam discretizes the search space. All displacements lie within some range, parametrized by parameter $r \in Z$, such that the displacement is the set of integer values in the range $[-r, r]$. By discretizing the search space, the optimization is able to take larger 'steps', whereas the in the continuous domain the optimizer would have to take, likely sub-voxel step sizes.

In order to achieve sub-voxel accuracies, ConvexAdam combines the discrete dual-optimization just described with continuous domain finetuning using gradient-based iteration as described by Equation (3). This finetuning uses the flow field generated by the discrete optimization. Instead of direct gradient descent, ConvexAdam uses an ADAM optimizer, which uses momentum to dampen oscillations and has been successful in optimizing deep neural networks in a variety of problem domains.

This combination of dual and discrete optimization with continuous domain ADAM optimization achieved state of the art performance in the large displacement dataset for MICCAI 2021 workshop, outperforming deep learning and conventional optimization methods.

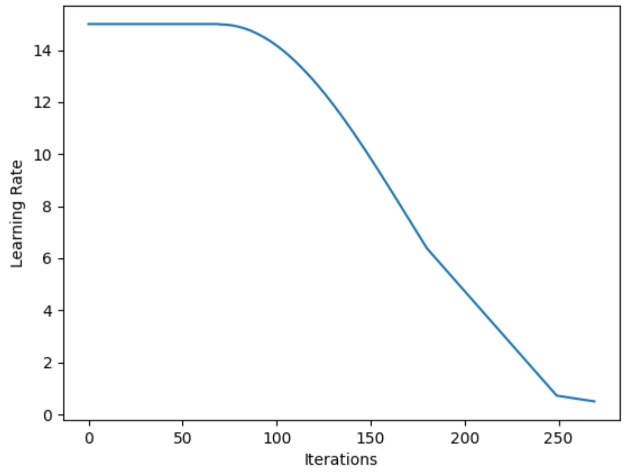

Fig. 1: Piece wise learning rate schedule. For step s, the learning rate schedule is constant for $s \leq 70$, cosine decay for $70 < s \leq 180$ and linear decay for $180 < s \leq 270$

3 Methods

3.1 Dataset

The NLST dataset contains 100 pairs of chest CT scans taken at patient inhalation and exhalation. The dataset also includes lung masks and corresponding pairs of keypoints between pairs of images. All volumes have been resampled onto an affine grid with dimensions (224,192,224) and isotropic spacing of 1.5mm. We further normalize the intensities of the volumes between between negative 4000 and positive 16000.

This dataset poses several challenges for registration. Firstly, there is large discrepancy in the volume of the lungs between inhalation and exhalation; this large displacement setting is one where deep learning methods have struggled to produce optimal results. Another challenge for deep learning methods is the size of the dataset is only 100, making it difficult to generalize to unseen data.

3.2 Approach

Given a pair of images, f and m, we implement continuous optimization of the flow field, using the iterative formulation described in Equations (1) and (3). Like ConvexAdam, we augment the gradients with momentum using the ADAM optimizer, which dampens the oscillations in the optimization trajectory achieved through unmodified gradient descent. Our similarity and regularization metrics are identical to the ones used by ConvexAdam. We briefly describe them here.

To compare similarity between the warped and target images, we compute the mean squared distance between the fixed and warped MIND features. This can be expressed as:

$$\mathcal{S} = \frac{1}{N}(F_f - F_m)^2 \tag{5}$$

where N is the size of the image in voxels, F_f is the MIND feature of the target image and F_m is the MIND feature of the warped image. The regularization function is simply the mean gradient of the displacement field:

$$\mathcal{R} = \frac{1}{N}|\nabla D| \tag{6}$$

Unlike ConvexAdam, which uses a constant learning rate of 1, we use a piecewise decreasing learning rate as shown in Figure 1. Intuitively, the large learning rate in the initial step enables us to take larger steps, while the smaller steps at later resolutions enables us to achieve sub-voxel displacements.

We also study the performance of the two optimization components (i.e. the discrete optimization and the continuous optimization) of ConvexAdam separately. This enables us to directly compare the continuous-domain optimization between the two methods and also provides novel insights into the performance of the discrete optimization.

We evaluate the performance of the registration using a keypoint displacement, which measures the distance (in voxels) of corresponding points in the target and warped images. A lower TRE score indicates better registration performance.

4 Results

Overall, the experiments conducted proved that IO could be greatly improved and that its value is not only limited to fine-tuning predictions but proves itself to be an outstanding standalone tool. The state-of-the-art performance observed in the experiments described below is achieved without requiring labels, keypoints, or segmentations to be provided to the algorithm besides the volume itself. It should be noted that for all experiments, all results converged well within the set number of iterations.

The coupled convex algorithm, ConvexAdam, and our algorithm result in an average TRE of 2.78mm, 1.14mm, and 0.888mm respectively.

The single LR of 1 for 100 iterations produces remarkably good registration results for such a simple solution, but with some fine-tuning can be improved greatly.

A very high learning rate of 15, allows all of the images to reach some local optima, with an average TRE of 2.44mm, but in the majority of cases this minimum is not a global minimum and for 100(%) of cases, sub-voxel accuracy cannot be achieved. However, images with large initial displacements significantly benefit from a high learning rate and are able to overcome local minima that were stifling performance when using an LR of 1.

In the hybrid cosine/linear decay LR schedule, 100(%) of paired images maintained or even exceeded their performance while those that were scoring above-voxel TRE improved most notably. Most notably image 0006 TRE improves from 16 to 1.48, achieving sub-voxel accuracy. The initial large learning rate of this schedule allows difficult local minima to be overcome while maintaining outstanding performance with the most basic optimizer.

This indicates that the network has the capability to find excellent solutions and local minima without the contributions of an initial prediction of a certain or any quality, eliminating the need for discrete optimization ahead of time.

Overall we observe excellent performance from IO alone without the need for an initial prediction. With the addition of a curated LR schedule, we see results even further improve to match the performance from when an initial prediction is set.

5 Conclusion

We show that our continuous domain optimization approach is able to beat state of the art performance of large displacement lung registration. The significance of this result is two-fold; first, the separation of ConvexAdam enables us to thoroughly examine the roles of the different optimization steps; such an examination was not present in the original ConvexAdam paper and is, therefore, provides greater insight into this method.

Next steps include testing the approach presented here against other medical image datasets. This should give us an idea of how generalizable our approach is to other anatomies and displacement settings. While the learning rate schedule was found through trial and error, future work can explore the possibility of a meta network to learn the parameters of the optimization, similar to hypermorph [4].

References

[1] M. Heinrich, M. Jenkinson, B. Papież, M. Brady, and J. Schnabel, “Towards Realtime Multimodal Fusion for Image-Guided Interventions Using Self-similarities,” *MICCAI*, 2013.

[2] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[3] H. Siebert, L. Hansen, and M. P. Heinrich, “Fast 3D Registration with Accurate Optimisation and Little Learning for Learn2Reg 2021,” in *Biomedical Image Registration, Domain*

Generalisation and Out-of-Distribution Analysis, M. Aubreville, D. Zimmerer, and M. Heinrich, Eds. Cham: Springer International Publishing, 2022, vol. 13166, pp. 174–179, series Title: Lecture Notes in Computer Science. [Online]. Available: https://link.springer.com/10.1007/978-3-030-97281-3_25

[4] A. Hoopes, M. Hoffmann, B. Fischl, J. Guttag, and A. V. Dalca, "HyperMorph: Amortized Hyperparameter Learning for Image Registration," *arXiv:2101.01035 [cs, eess]*, May 2021, arXiv: 2101.01035. [Online]. Available: http://arxiv.org/abs/2101.01035