
Fast Learning-based Registration of Sparse Clinical Images

Kathleen M. Lewis
MIT

Guha Balakrishnan
MIT

Natalia S. Rost
MGH

John Guttag
MIT

Adrian V. Dalca
MIT, MGH

Abstract

Deformable registration of clinical scans is a fundamental task for many applications, such as population studies or the monitoring of long-term disease progression in individual patients. This task is challenging because, in contrast to high-resolution research-quality scans, clinical images are often sparse, missing up to 85% of the slices in comparison. Furthermore, the anatomy in the acquired slices is not consistent across scans because of variations in patient orientation with respect to the scanner. In this work, we introduce Sparse VoxelMorph (SparseVM), which adapts a state-of-the-art learning-based registration method to improve the registration of sparse clinical images. SparseVM is a fast, unsupervised method that weights voxel contributions to registration in proportion to confidence in the voxels. This leads to improved registration performance on volumes with voxels of varying reliability, such as interpolated clinical scans. SparseVM registers 3D scans in under a second on the GPU, which is orders of magnitudes faster than the best performing clinical registration methods, while still achieving comparable accuracy. Because of its short runtimes and accurate behavior, SparseVM can enable clinical analyses not previously possible. The code is publicly available at voxelmorph.mit.edu.

1 Introduction

Deformable registration establishes a dense, non-linear correspondence between a pair of 3D MRI scans. Deformable registration is a fundamental step in medical image analyses, such as population studies and disease monitoring in individual patients. For example, when clinicians are monitoring a brain lesion in a patient, it is helpful to register longitudinal scans to get an accurate understanding of the disease progression, which in turn leads to affecting treatment. Since registration is a difficult task, it has been an active area of research for decades.

Most of the work in this field focuses on high-resolution research-quality scans. However, in clinical settings, scanning time is often limited. For 3D imaging modalities such as MRI, this often means few 2D slices are acquired instead of a full 3D volume, leading to spatially sparse scans. Each 2D slice is high resolution, but the 3D volume can be missing up to 85% of the slices compared to a research-quality scan, where the acquired slices are often as dense as the in-plane resolution. The wide spacing between slices causes drastic discontinuities in anatomy between neighboring slices (Fig. 1, left). Furthermore, the anatomy captured by the slices is not consistent across different images. Depending on the orientation of the patient with respect to the scanner, different anatomy may or may not be present in the obtained slices. Full 3D registration of these sparse images presents technical challenges because most current registration methods use image gradients, which require spatial continuity.

One approach to registering sparse scans is to first estimate the missing data, which is most often achieved by linear interpolation. This results in full resolution images, but can be inaccurate. The linearly interpolated images are then registered using an existing registration method.

Most classical registration methods optimize over the space of displacement vector fields, including elastic-type models [4, 18], statistical parametric mapping [2], free-form deformations with b-splines [17], and Demons [21]. A recent patch-based method [9] builds on discrete registration methods and explicitly accounts for sparsity in clinical data. Since these classical methods require solving a high dimensional optimization on each new pair of images, their runtimes are slow.

Recent methods use neural networks to learn a function that takes in two images as input and outputs a deformation field. Many are supervised, requiring ground truth warp fields or segmentations [13, 15, 19, 22]. Early unsupervised, or end-to-end, methods [10, 14] were only evaluated on a limited subset of volumes, such as 2D slices. A very recent method, VoxelMorph (VM) [5, 6, 8], develops an unsupervised method that is evaluated on 3D MRI scans, but these scans are of research quality and high resolution. As we demonstrate in our experiments, it under-performs on clinical images. In this work, we build upon VoxelMorph, to accurately register both sparse clinical scans and research scans, without losing the rapid runtime demonstrated by VM.

We evaluate our method, which we call Sparse VoxelMorph (SparseVM), on a clinical study of stroke patients [16] containing T2-FLAIR MR brain scans. We demonstrate that our method is 1) more than $100\times$ faster than the best classical methods, while maintaining high accuracy, and 2) more accurate than the best learning-based methods for registering clinical images.

2 Method

Let f, m be two volumes defined over a n -D spatial domain $\Omega \subset \mathbb{R}^n$ (where $n=3$ in this paper). We focus on atlas-based registration, in which an atlas, or a reference volume, is registered to each clinical subject scan; this is commonly used in population analysis [20]. Therefore, we define f as a linearly-interpolated clinical subject scan and m as a full resolution atlas. For simplicity, we assume f and m contain single-channel, grayscale data and are affinely aligned as a preprocessing step, so that the images only contain non-rigid misalignments.

SparseVM adapts the fast, learning-based method VoxelMorph to accurately register clinical images. VoxelMorph is an unsupervised method that uses a convolutional neural network to learn a function $g_\theta(f, m) = \phi$ to compute a deformation field, ϕ , for a pair of given images. At test time, VoxelMorph evaluates this function on the two input images, and outputs the deformation field, ϕ , as well as the registered image, $m \circ \phi$ (m warped by ϕ). We use the same architecture proposed in the original VoxelMorph paper [5].

VoxelMorph was demonstrated on high resolution images, in which all voxels are equally reliable, and therefore contribute equally to the loss function during training. When used with clinical images which contains linearly interpolated voxels, VoxelMorph under-performs, likely because it doesn't differentiate between interpolated and acquired voxels. This discrepancy motivates our method, which differentiates interpolated voxels from acquired voxels.

2.1 Loss Functions

The VoxelMorph loss function is based on that used in many classical registration methods, balancing an image matching term with a regularization or smoothness term. The similarity loss, L_{sim} , measures the similarity in appearance between f and $m \circ \phi$. The smoothness loss function, L_{smooth} , encourages a smooth displacement field, ϕ , to make sure the transformation, $m \circ \phi$, is physically realistic. The unsupervised loss is then:

$$L_{VM}(f, m, \phi) = L_{sim}(f, m \circ \phi) + \lambda L_{smooth}(\phi), \quad (1)$$

where λ is a regularization parameter. In VoxelMorph, L_{sim} is implemented as either mean squared error (MSE) or cross correlation. L_{smooth} penalizes the local spatial gradients of u , ∇u , where u is the displacement such that $\phi = Id + u$ and Id is the identity transform:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p)\|^2. \quad (2)$$

We propose a new image similarity loss that allows VM to perform accurate registration on both clinical and high resolution scans. In VoxelMorph, L_{sim} equally weights all voxels. SparseVM

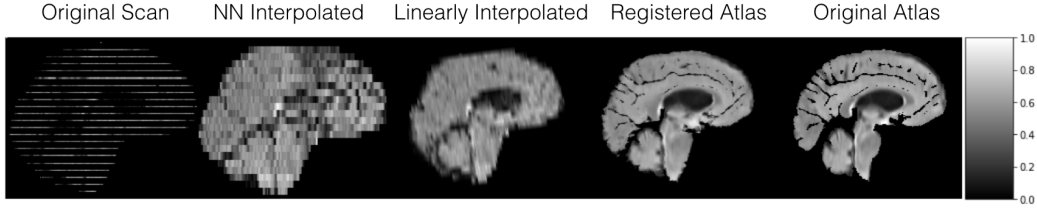


Figure 1: Example sagittal slices of a stroke subject and atlas image pair. The original subject scan has approximately 14% of the slices in the full resolution atlas. Shown from left to right are the original subject slices, nearest neighbor interpolated subject image, linearly interpolated subject image, atlas registered by sparseVM, and the original atlas. Despite significant sparsity in the subject image, SparseVM is able to accurately register the atlas.

introduces the concept of weighting voxels in proportion to their reliability. The similarity loss, L_{sim} , is modified to take in masks, or per pixel weightings. In the case of atlas-based registration, the moving image, M , is a full-resolution image with equal confidence across all pixels. In each clinical scan, F , however, the majority of the full resolution volume contains unobserved voxels. Therefore, we focus only on the fixed image mask, M_f , that indicates which voxels in the subject image f are interpolated and which voxels originate from the clinical image slices.

The values of this mask are initially continuous after interpolation during the affine alignment in pre-processing. The continuous values can be interpreted as voxel weightings. Intuitively, voxels with low values are not likely to contain information from the acquired slices. SparseVM evaluates the mean squared error on these voxels, which are more likely to contain information from the acquired scanner slices:

$$L_{sim}(f, M_f, m \circ \phi) = \frac{1}{|M_{f'}|} \sum_{i \in M_{f'}} (f_i - (m \circ \phi)_i)^2, \quad (3)$$

where $M_{f'}$ is the set of locations of nonzero voxels in M_f . By only penalizing the voxels from the original slices, SparseVM prioritizes aligning true voxels over possibly-misinformed interpolated voxels. VoxelMorph can be interpreted as a special case of SparseVM, in which the mask is 1 for all voxels.

2.2 Implementation

We implemented SparseVM using Keras [7] with a Tensorflow backend [1]. We use the ADAM optimizer [12] with a learning rate of $5e^{-4}$. We find the optimal spatial smoothness regularization parameter λ by performing a grid search and found the optimal value to be 0.4.

3 Experiments

3.1 Dataset

We demonstrate our method on a clinically acquired stroke dataset, which contains manual segmentations of the ventricles. Figure 1 shows the original subject scan, nearest neighbor interpolated subject image, and linearly interpolated subject image for an example stroke subject, as well as the original atlas and the atlas registered by SparseVM. The atlas is a previously-built full resolution scan [20], while the stroke subject scans have approximately 14% of the number of slices in the atlas. The figure shows 2D sagittal views, however, all experiments in this paper are on 3D volumes. During pre-processing, the subject images were linearly interpolated to full resolution affinely aligned to the T2-FLAIR atlas. The images were padded with zeros to be of size 160x192x160.

3.2 Baseline Methods

We compare against the patch-based registration (PBR) method [9] because it is a recent method that shows promising results for clinical image registration. We use the parameters shown to be optimal in the original paper. We also compare to the ANTs registration algorithm [3] with the optimal parameters from the PBR paper. Finally, we compare against VoxelMorph [5] with an MSE loss.

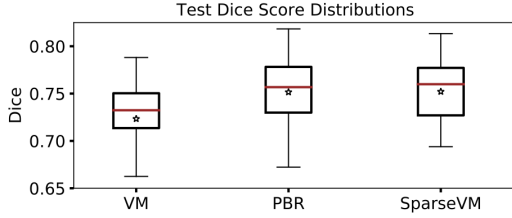


Figure 2: Boxplot of Dice scores for VoxelMorph, SparseVM, and PBR on the test subjects. For each, the red line indicates the median, and the star indicates the mean. The box edges show the 25 and 75 percentiles. The whiskers show the 10 and 90 percentiles. SparseVM is significantly better than VM and is comparable with PBR while requiring significantly less time.

Method	Average (SD)	Median (MAD)	GPU (sec)	CPU (sec)
Affine only	0.621 (0.072)	0.623 (0.061)	0	0
ANTs (optimized)	0.722 (0.031)	0.730 (0.023)	-	9059 (2023)
PBR (best)	0.752 (0.037)	0.757 (0.027)	-	9269 (5134)
VoxelMorph	0.723 (0.045)	0.732 (0.032)	0.313 (0.046)	40 (0.693)
SparseVM	0.752 (0.036)	0.760 (0.028)	0.303 (0.047)	41 (0.584)

Table 1: Average Dice scores and test runtimes for affine alignment, ANTs, PBR, VoxelMorph, and SparseVM. The average Dice score and median Dice score are computed over all test subjects on the left and right ventricle labels. The standard deviation and median absolute deviation are shown in parentheses. The runtimes are calculated after pre-processing. SparseVM is as accurate as the most accurate baseline, PBR, and orders of magnitude faster.

3.3 Evaluation

We evaluate our method and the baseline methods using the Dice metric [11] for ventricles (the only structure for which labels were available). Dice measures the volume overlap of anatomical segmentations, with 0 representing no overlap and 1 representing perfect overlap. We select only subjects that have good affine registration (as measured by Dice score) in order to separate out effects that are not meant to be captured by SparseVM or any of the baseline methods. In order to compare our method with the current state of the art, we evaluate all methods on 51 of the test subjects from [9], which were randomly chosen in the original study.

3.4 Results

Table 1 summarizes the average and median dice scores for all subjects on the ventricle structures and the average test runtimes of the methods. SparseVM yields higher Dice scores than both ANTs and VM on 92% of subjects and higher Dice scores than PBR on 53% of subjects. SparseVM requires orders of magnitudes less runtime than ANTs and PBR and has a runtime on par with VoxelMorph on a single-threaded CPU. There are no GPU implementations of ANTs or PBR. However, we expect these algorithms to still be much slower than SparseVM on the GPU, since SparseVM only requires evaluating a function at test time, while the other methods are iterative. Both ANTs and PBR take on the order of two hours on the CPU, while SparseVM takes less than a minute on the CPU and less than a second on the GPU. Figure 2 is a boxplot of the test Dice scores for the three most accurate methods. The 25th and 75th percentiles of SparseVM are much higher than those of VM and approximately the same as those of PBR.

4 Conclusion

We introduce Sparse VoxelMorph, which is a fast, unsupervised method that generalizes the state-of-the-art VoxelMorph method to sparse, low-resolution clinical image registration. SparseVM improves performance accuracy over VoxelMorph while maintaining comparable speed. It is orders of magnitude faster than methods that achieve state-of-the-art accuracy for clinical image registration, while achieving comparable or better Dice scores. Since SparseVM can accurately register pairs of clinical images in under a second on the GPU, it enables clinical image analyses that were not previously possible. It is actively being translated to be used with large clinical collections.

References

- [1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [2] J. Ashburner and K. Friston. Voxel-based morphometry-the methods. *Neuroimage*, 11:805–821, 2000.
- [3] B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
- [4] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46:1–21, 1989.
- [5] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *arXiv preprint arXiv:1809.05231*, 2018.
- [7] F. Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- [8] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. volume 11070, pages 729–738. *MICCAI: Int. Conf. on Medical Image Computing and Computer Assisted Intervention.*, LNCS, 2018.
- [9] A. V. Dalca, A. Bobu, N. S. Rost, and P. Golland. Patch-based discrete registration of clinical brain images. pages 60–67. *Patch-Based Techniques in Medical Imaging: Second International Workshop, Patch-MI 2016, Held in Conjunction with MICCAI 2016*, 2016.
- [10] B. de Vos, F. F. Berendsen, V. M. A., M. Staring, and I. Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212, 2017.
- [11] L. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [12] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [13] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen. Robust non-rigid registration through agent-based action learning. pages 344–352. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2017.
- [14] H. Li and Y. Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv*, 2017.
- [15] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec. Svf-net: Learning deformable image registration using shape matching. pages 266–274. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2017.
- [16] N. Rost, K. Fitzpatrick, A. Biffi, A. Kanakis, W. Devan, C. D. Anderson, L. Cortellini, K. L. Furie, and J. Rosand. White matter hyperintensity burden and susceptibility to cerebral ischemia. *Stroke*, 42(12):2807–2811, 2010.
- [17] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformation: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
- [18] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.
- [19] H. Sokooti, B. de Vos, F. Berendsen, B. P. Leveveldt, I. Išgum, and M. Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. pages 232–239. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2017.

- [20] R. Sridharan, A. V. Dalca, K. M. Fitzpatrick, L. Cloonan, A. Kanakis, O. Wu, K. L. Furie, J. Rosand, N. S. Rost, , and P. Golland. Quantification and analysis of large multimodal clinical image studies: Application to stroke. pages 18–30. International Workshop on Multimodal Brain Image Analysis, Springer, 2013.
- [21] J. Thirion. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.
- [22] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver: Fast predictive image registration- a deep learning approach. *NeuroImage*, 158:378–396, 2017.