CNN Driven Sparse Multi-Level B-spline Image Registration

Pingge Jiang James A. Shackleford
Department of Electrical and Computer Engineering
Drexel University
{pingge,shack}@drexel.edu

Abstract

Traditional single-grid and pyramidal B-spline parameterizations used in deformable image registration require users to specify control point spacing configurations capable of accurately capturing both global and complex local deformations. In many cases, such grid configurations are non-obvious and largely selected based on user experience. Recent regularization methods imposing sparsity upon the B-spline coefficients throughout simultaneous multi-grid optimization, however, have provided a promising means of determining suitable configurations automatically. Unfortunately, imposing sparsity on over-parameterized B-spline models is computationally expensive and introduces additional difficulties such as undesirable local minima in the B-spline coefficient optimization process. To overcome these difficulties in determining B-spline grid configurations, this paper investigates the use of convolutional neural networks (CNNs) to learn and infer expressive sparse multi-grid configurations prior to B-spline coefficient optimization. Experimental results show that multi-grid configurations produced in this fashion using our CNN based approach provide registration quality comparable to L_1-norm constrained over-parameterizations in terms of exactness, while exhibiting significantly reduced computational requirements.

1. Introduction

Deformable image registration is a task that attempts to precisely reproduce the spatial transform which maps two images collected at different points in time or across different modalities into a common coordinate system. The computation of such spatial transforms is inherently ill-posed; however, the value of recovering such transforms to the medical community alone, as they relate to quantifying anatomical motion, has motivated extensive formalization of the problem over the past two decades. This pursuit has led to the investigation of deformation models such as thin-plate splines [2], B-splines [18], optical flow [26], and contour driven methods [11]. The ill-posed nature of the problem has driven the development of representative objective functions for parameter optimization [24, 6, 29, 28], and the complexity of the problem has led to improvements in accuracy and computational efficiency [20, 14, 21] in the interest of improving viability for routine clinical use.

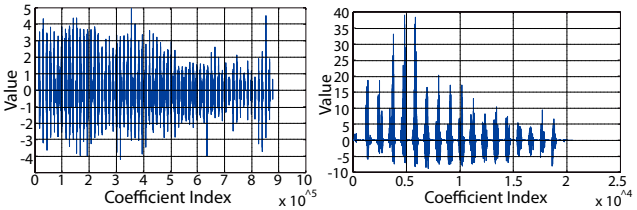
Due to their robustness and suitability for performing multi-modal registration, B-spline based models have been widely investigated [17, 16, 9, 28, 25, 20, 5]. In all cases, the selection of a B-spline control grid configuration capable of expressing the desired underlying transformation is an integral aspect of these methods, which ultimately relies upon the degree and nature of the underlying image deformation attempting to be recovered. Consequently, a control grid configuration capable of parameterizing the transform is generally selected manually by the operator after inspection of the two images undergoing registration. Fine grids with closely separated control points are ideal for enabling the expression of complex local deformations whereas coarse grids with further separation are more suitable for recovering larger, gradual deformations. The use of a single fine grid providing an adequately high degree of freedom to express the underlying transform would be ideal, however the deviation of current similarity metrics from the true, and unknown, underlying objective function coupled with the over parameterization provided by such a grid of fine spatial resolution results in a highly non-convex B-spline coefficient optimization process that is susceptible to being caught in false local minima, resulting in poor transform recovery.

Pyramidal B-spline registration [10, 19, 27, 30, 15] is a commonly employed approach to addressing this issue. In such a scheme, the process begins with performing parameter optimization using a coarse grid. Once the stopping criteria is met the coarse grid is discarded, a finer grid is employed, the parameters obtained from the coarser grid optimization are fit to the new finer grid, and the optimization process continues. This process is continued over several levels of increasing grid resolution until the operator is satisfied. The motivation driving the use of such pyramidal

grid methods is the assumption that all grids exhibit gradients that drive the optimization towards the proximity of the desired objective function minima under the finest control grid configuration, and although additional false local minima are introduced with each increase in grid resolution, the optimization process is unlikely to arrive at such solutions due to being driven away from them through the prior optimization of the coarser grids in the pyramid. In such a fashion, large sweeping deformations are recovered early in the pyramid and complex local deformations are recovered in the later, finer levels of the pyramid. However, as each layer in the pyramid is optimized independently, image regions exhibiting large sweeping motions that are adequately captured at the coarser levels become over parameterized at later levels in the pyramid. For such regions, which are generally larger regions of relatively uniform intensity, this increase in parameterization can result in the optimizer producing physically unrealistic solutions due to the decreased spatial extent of the B-spline basis local support, which imposes first order continuity in the resulting transform. While this decoupling of adjacent regions is desirable for image regions exhibiting complex local deformations, it is equally undesirable for regions lacking in such complexity. As a result, the pyramidal B-spline grid configuration designed by the operator is vitally important in producing transforms adequately representative of the deformation that occurred between the two images involved in the registration process.

The over parameterization of the underlying transform provided by any given B-spline control grid is evidenced by inspecting the sparsity of the B-spline basis coefficient values produced by the optimizer for a successful registration. Figure 1 shows coefficient values for both fine and coarse control grids after arriving at physically meaningful registrations via optimization. Demonstrably, many of these coefficients are found to be superfluous, taking on values at or near zero. Consequently, Shi *et al*. [23, 22] have investigated the simultaneous optimization of multi-level B-spline grids, coupling them through the L_1-norm sparsity constraint. Inspection of the resulting B-spline grids reveals large sweeping deformations being parameterized by the coarser level grids and local complex deformations being expressed by the finer grids. Furthermore, noise is suppressed in the resulting transforms due to the imposition of sparsity. Pragmatically, however, the simultaneous optimization of multiple grids is a slow and memory intensive process, which is further encumbered by the evaluation of the L_1-norm.

In this paper, we introduce a convolutional neural network (CNN) architecture capable of learning and later inferring the sparsity of multi-level B-spline grid configurations based on features present in the input image set prior to parameter optimization. The simultaneous optimization of multiple grids under the L_1-norm can be viewed as a means



(a) (b)

Figure 1: Coefficient values for both (a) fine and (b) coarse control grids after optimization. Many coefficients are at or near zero, demonstrating sparsity.

of simultaneously determining zero value control points while solving for optimal B-spline basis coefficients. The approach detailed in this paper divides this process into two distinct steps. Control points ill-suited for parameterization of the deformation transform under recovery are learned by a CNN using training data generated from L_1-norm constrained multi-level B-spline grid registrations. Once such a CNN is trained to recognize the support required to express the deformation across all regions of the image, control point coefficients deemed superfluous by the CNN are constrained to zero while remaining coefficients are optimized to arrive at the transform best describing the image deformation. Experimental results show approximately a 90% reduction in optimization parameters while overall registration quality is improved.

2. Related work

The use of multiple grids of varying spatial resolution has been throughly investigated due to the sensitivity of the optimization process to local minima while attempting to accurately recover complex local deformations. Andronache *et al*. [1] introduce a non-rigid registration algorithm that classifies sub-image consistency using the Moran information consistency test, the results of which are used to drive a hierarchical subdivision procedure. A different similarity metric is then used for each resulting level in order to provide robust and computationally efficient matching between corresponding image regions. Buerger *et al*. [3] propose a method that adaptively sub-divides image regions based on the presence of image features and motion complexity. Image regions exhibiting similarity within these criteria are coupled into single registration components. Work by Jiang [8] introduces an octree representation for groups of aligned B-spline control grids providing subdivided support regions when ordered from coarse to fine. Heuristic features from image pairs undergoing registration are used to construct a pruned octree representing an effectively non-uniform grid deemed best suited for recovering the image deformation. Shi *et al*. [23, 22] introduce a multi-level B-

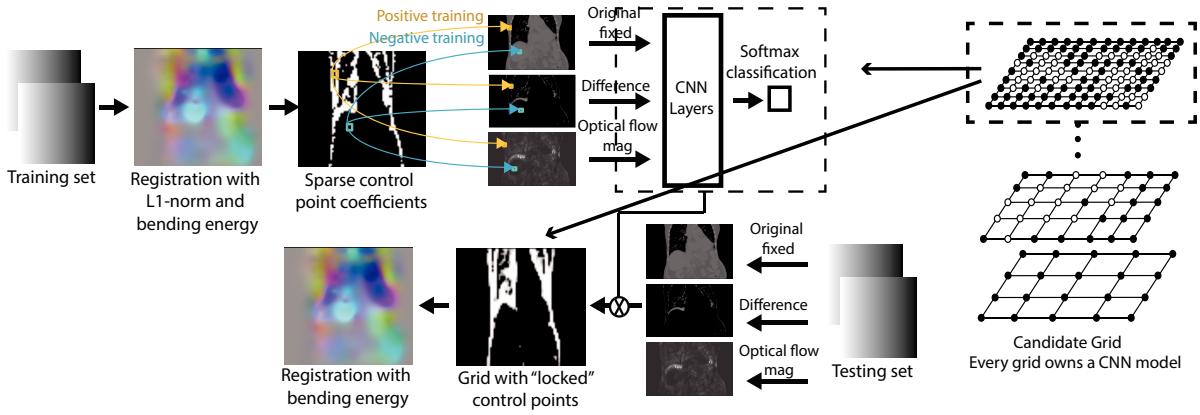


Figure 2: Complete workflow. The training process consists of using pyramidal B-splines regularized by the L_1-norm to generate ground truth training data. Preprocessing is employed to generate input channels corresponding to grid regions within the pyramid. The preprocessing for testing data is performed similarly with the resulting grid configurations produced by the network being subjected to a non-regularized optimization process.

spline parameterization where multiple grids of increasing spatial resolution are optimized simultaneously. The solution is regularized by the addition of bending energy and parameter sparsity penalties to the common sum of squared differences (SSD) similarity metric. The use of the sparsity metric, in this case, forgoes the need for feature driven grid level assignment of image sub-regions as it is determined automatically through the optimization process.

Most recently, CNN architectures have begun to inspire new approaches to image registration. Miao *et al*. [13] use a regression model that recursively estimates rigid transformation parameters for image regions. Instead of optimizing parametric transformation model parameters based on a cost function, the proposed method uses CNN regressors to extract image features, which are used to predict transformation parameters for image correspondence. Liao *et al*. [12] repose the registration problem as a series of actions that can be performed by an entity they deem the agent. This agent consists of two CNNs, the first of which possesses a large field of view (FOV) with a focus on obtaining global anatomical understanding and a coarse global alignment. The second CNN has a more limited FOV and is tasked with performing finer local alignment.

In this paper, we propose a CNN driven multi-grid B-spline method that focuses on learning the most suitable parameterization for describing the deformable transform between two images prior to the parameter optimization process. To this end, the proposed CNN is trained using deformable registrations normalized by the L_1-norm in a fashion similar to that performed by Shi *et al*. [22]. Once trained, we demonstrate that multi-grid configurations possessing far fewer parameters are generated by the CNN given an input image pair without sacrificing registration quality while dramatically reducing the time and memory required by the subsequent optimization process.

3. L_1-norm Regularized Registration

Here we briefly describe the L_1-norm regularized registration process introduced by Shi *et al*. [22], which is employed to generate data referenced throughout the remainder of this paper as both training data and as a basis for results comparison. Given a three dimensional thoracic CT image F and a corresponding CT image M acquired at different respiratory phases, the non-rigid B-spline registration aims to recover the deformation field $\vec{v}$ that most accurately maps voxels in M into the coordinate system of F. This is accomplished by optimizing a cost function C such that:

$$\vec{v}* \leftarrow \arg\min_{\vec{P}} C\left(F, M, \vec{P}\right) \tag{1}$$

where $\vec{P}$ represents B-spline coefficients to be optimized and $\vec{v}*$ is the resulting vector field, parameterized by $\vec{P}$, mapping M into F. The total cost function C consists of three additive terms in order to recover a globally acceptable deformation field: a similarity term measuring the voxel-wise sum of squared difference (SSD) between F and M, denoted as E_S; a bending energy penalty enforcing second-order smoothness in $\vec{v}*$, denoted as E_R; and the L_1-norm penalty enforcing coupled multi-grid sparsity. The complete cost function C can be expressed as [23]:

$$C = E_S(F, M, \vec{P}) + \lambda_R \sum E_R(\vec{P}) + \lambda_S \left\|\vec{P}\right\|_1 \tag{2}$$

where λ_R and λ_S weigh the smoothness and sparsity penalties, respectively. As illustrated by Figure 3, the inclusion of the L_1-norm penalty serves to provide coupling between

the various grid layers while enforcing a degree of noise suppression in the resulting deformation vector field. As a consequence of this coupling, subsets of $\vec{P}$ providing support over a given subregion of the image will be largely constrained to a single grid level having the greatest suitability for expressing the deformation field in that subregion. In the case of the uniform cubic B-spline basis, every voxel in the N-dimensional image volume is supported by the tensor product of the four neighboring control points in each of the N grid dimensions. The deformation field $\vec{v}$ is parameterized by B-spline basis coefficients $\vec{P}$, which provides a sparse representation of $\vec{v}$ while additionally introducing first order continuity. The dense, voxel-wise deformation field of $\vec{v}$ at each level L may be interpolated using the B-spline basis and the set $\vec{P}$ such that:

$$v_L(x) = \sum_{(l,m,n)=0}^{3,3,3} B_l(u)\, B_m(v)\, B_n(w)\, P_{i+l,\, j+m,\, k+n} \tag{3}$$

in the x-direction, and similarity for y and z directions. Here, (i, j, k) denote the coordinates of a given voxel in F, (u, v, w) are the local coordinates of the voxel within its housing support region normalized within $[0, 1]$ used for evaluating the B-spline basis function, and B is the uniform cubic B-spline basis function. For a multi-level scheme consisting of M grids, the final deformation field is the summation of local deformation fields from each grid level:

$$\vec{v}* = \sum_{L=1}^{M} \vec{v}_L \tag{4}$$

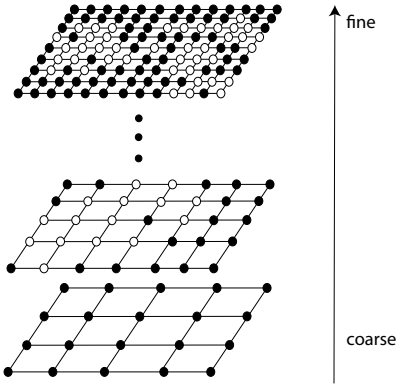


Figure 3: Multi-level 2D B-spline grid. The L_1-norm sparsity constraint couples the various levels, resulting in levels best suited for parameterizing a given local set of deformation vectors having non-zero control points (black) while control points providing support in less suitable levels are driven to zero (white) through the optimization process.

4. Learning Grid Configurations with CNNs

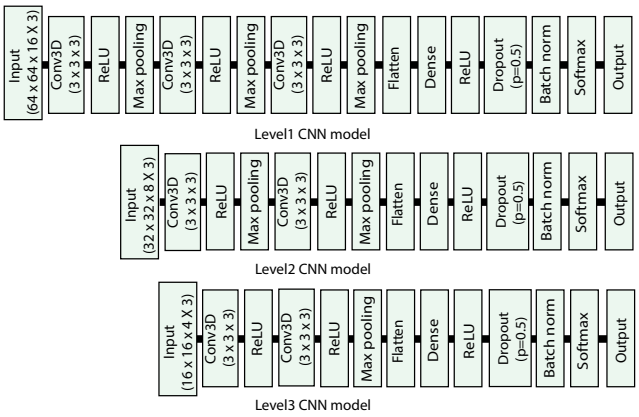


Figure 4: CNN architecture for each B-spline control grid level. The CNN for level 1 accepts 3 input channels of dimensions $64 \times 64 \times 16$. The CNN for level 2 accepts 3 input channels of of $32 \times 32 \times 8$. The CNN for level 3 accepts 3 input channels of $16 \times 16 \times 4$. Fewer layers are used for finer grids due to the decreased field of view within the smaller input patches.

For a given pair of image volumes to undergo deformable registration, the proposed method, as illustrated in Figure 2, aims to accurately predict the zero valued coefficients in an over parameterized multi-level uniform cubic B-spline grid configuration. To this end, a CNN architecture is trained using the B-spline control point coefficients and grid configurations produced by L_1-norm regularized registrations. In this particular study, we use 4D CT thoracic image volume sets capturing the respiration cycle. Each 4D CT dataset consists of ten 3D CT volumes, where each 3D volume captures the thorax at equally spaced phases throughout one full respiration cycle. We perform L_1-norm regularized registration between image volumes at the extrema of the respiration cycle: initial full-exhale to full-inhale and again between full-inhale to final full-exhale; thereby using only three of the full ten phases available to produce two registrations. The resulting sparse B-spline control grids consist largely of zero valued control point coefficients, which we will call "locked control points," as well as a minority of non-zero valued coefficients, which we will call "free control points." These "free" and "locked" designations for control points form the class labels for the CNN training data. Preprocessing is performed on the associated image volumes to produce feature channels that will serve as inputs to the CNN input layer. One CNN is trained per B-spline control grid layer. For example, the CNN architecture for the three layer grids experimentally validated in this paper is illustrated in Figure 4. Once trained, unregistered image pairs undergo preprocessing, are fed into each CNN,

a multi-level grid with classified control points is produced, and the deformation field $\vec{v}*$ between the two images is produced by optimizing "free" control point coefficients while "locked" coefficients are held at zero.

4.1. Preprocessing & CNN Inputs

Input channels to the proposed CNN architecture consist of the unaltered fixed image F as well as two other input channels derived from the fixed and the moving images in combination. These two additional inputs are generated in order to expose meaningful information indicating the severity and nature of the deformation between F and M to the convolution layers of the CNN. Specifically, three input channels are provided to the input layer of the CNN: the unaltered fixed image F, an estimation of optical flow magnitude, and a difference image. The estimation of optical flow magnitude is generated by running the Lucas-Kanada method using F and M for a single iteration. The difference image is simply the result of subtracting the fixed image from the moving image, $F - M$.

4.2. Generation of Training Data

Multi-layer B-spline control grids with control points labeled as either "free" or "locked" are generated by performing B-spline registration using the SSD similarly metric penalized by bending energy ($\lambda_R = 0.001$) and the L_1-norm ($\lambda_S = 0.04$). Three grid levels of increasing control point resolution are used with each possessing an even support region subdivision of the previous coarser level. Specifically, Level 1 consists of $35 \times 35 \times 35$ control points, Level 2 consists of $67 \times 67 \times 67$ control points, and Level 3 consists of $131 \times 131 \times 131$ control points. For the $512 \times 512 \times 128$ voxel input images undergoing registration in this study, this translates to support regions of $16 \times 16 \times 4$, $8 \times 8 \times 2$, and $4 \times 4 \times 1$ voxels for Levels 1, 2, and 3, respectively.

Control point coefficient values $\vec{P}*$ minimizing (2) are estimated via quasi-Newtonian optimization (L-BFGS-B). The efficiency of the cost function gradient $\partial C/\partial \vec{P}$ calculation is markedly accelerated by employing the map-reduce framework proposed by Jiang [7], which eliminates unnecessary redundant loads of intermediate voxel-wise $\partial C/\partial \vec{v}$ calculations by generating so called "Z-values":

$$Z_{rgn,l,m,n} = \sum_{(z,x,y)=0}^{N_{z,x,y}} \frac{\partial \mathrm{C}}{\partial v\,(x,\ y,\ z)} B_l\,(u)\,B_m\,(v)\,B_n\,(w) \tag{5}$$

which weight and sum each voxel-wise $\partial C/\partial \vec{v}$ value within a local support region by all 64 possible piecewise B-spline basis function weighting combinations; thereby resulting in 64 "Z-values" per local support region. The computation of the cost function gradient $\partial C/\partial \vec{P}$ for any given control point is subsequently achieved by mapping the 64 "Z-values" corresponding to the control point and reducing them via summation to a single value.

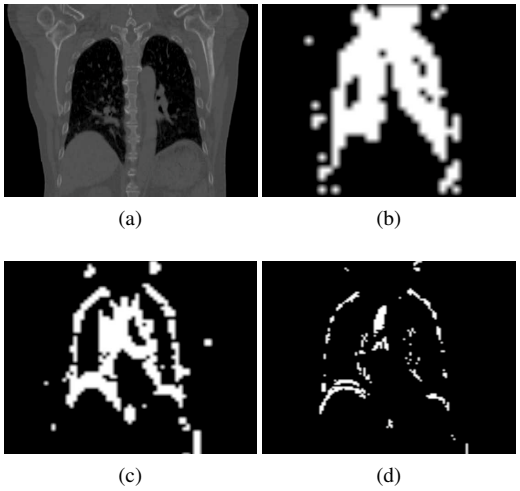


Figure 5: Visualization of control point labels produced by a typical L_1-norm regularized registration. (a) Coronal view of the $512 \times 512 \times 128$ voxel fixed CT image. (b-d) Control grid slices visualizing "free" control points in white and "locked" control point in black for the (b) Level 1: $35 \times 35 \times 35$ control point grid, (c) Level 2: $67 \times 67 \times 67$ control point grid, and (d) Level 3: $131 \times 131 \times 131$ control point grid.

Once the registration process has completed, control points possessing zero valued B-spline coefficients are considered "locked" and all other control points are considered "free." Figure 5 visualizes a typical example of "locked" and "free' control points produced by the registration process. CNN training patches are next extracted from the three preprocessed input channels, namely the unaltered fixed image F, the estimation of the optical flow magnitude, and the difference image $F - M$. One training patch for each input channel is extracted per labeled control point produced by the L_1-norm regularized registration process. As illustrated in Figure 6, each extracted training patch is centered about a control point, has dimensions corresponding to the extent of the control point's local support in the voxel coordinate system, and is labeled either "free" or "locked" in correspondence with the control point at its center.

5. Experimental Results

The experiments detailed in this section were performed on a machine equipped with dual HyperThreaded 3.2 GHz Intel Xeon OctoCore E5-2630v3 processors, 512 GB DDR4 RAM, and dual NVIDIA GeForce GTX 980 GPUs. Tensorflow was used as the backend for the neural network architecture.

Here we investigate the performance of the proposed al-

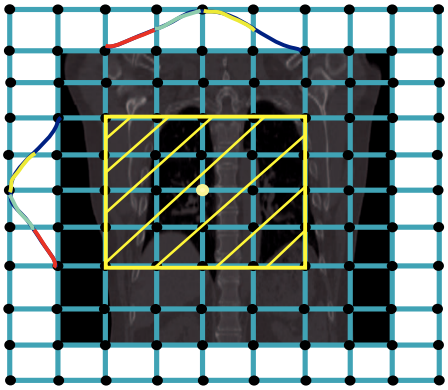


Figure 6: Training patch extraction. A uniform cubic B-spline grid superimposed upon a coronal CT image slice. Basis functions shown in periphery to demonstrate local support. The local support region for the control point in yellow is illustrated by the surrounding yellow box. This sub-image within the box is extracted as a training patch.

gorithm for a three level B-spline grid where Level 1 consists of $35 \times 35 \times 35$ control points, Level 2 consists of $67 \times 67 \times 67$, and Level 3 consists of $131 \times 131 \times 131$ control points. The image volume data used in the training and validation consists of 4D thoracic datasets, each capturing a single full respiration phase. This data was obtained from two different sources. The first source is the DIR-Lab 4D CT lung dataset [4], which we will refer to as "Set 01." This is supplemented by anonymized thoracic 4D CT image sets provided by clinical collaborators, which we will refer to as "Set 02." Set 01 provides 10 pairs of inhale and exhale image volumes from 5 different patients having axial slice dimensions of 512×512 voxels with the number of axial slices varying between 120 and 136 due to acquisition inconsistencies. Set 02 provides 12 pairs of inhale and exhale images from 6 different patients having axial slice dimensions of 512×512 voxels with the number of axial slices varying from 128 to 152, again, due to acquisition inconsistency. Image volumes were resampled uniformly to $512 \times 512 \times 128$ voxels with an intervoxel spacing of $1.16mm \times 1.16mm \times 2.5mm$. The total data available across the two sets provides 943,250 control point centric patches in grid Level 1; 6,616,786 patches in grid Level 2; and 49,458,002 patches in grid Level 3.

Control point centric image patches extracted from four pairs of images from Set 01 and four pairs of images from Set 02 are used as CNN training data. Due to the high degree of sparsity exhibited by control points optimized under the L_1-norm, the number of available training patches corresponding to the "locked" class is much greater than the number of training patches in the "free" class—this extreme imbalance can be seen in Figure 5. Consequently, the number of training patches corresponding to the "locked" class are reduced to meet that of the "free" class. A random sampling of training patches from the "locked" class was therefore used to balance the number of positive and negative examples in an unbiased fashion. In total, the number of used training patches for grid Level 1 CNN is 64,564 (32,282 positive and 32,282 negative); 233,250 for grid Level 2 (again, evenly balanced); and 679,256 for grid Level 3 (evenly balanced). Patches from other images are used for testing only without random sampling.

The L_1-norm registration used to obtain control point grid class labels was ran for 20 L-BFGS-B optimization iterations. The vector fields produced by this same registration is also later used as a basis for comparing our algorithm's performance. Neural network models are trained for 20 epoches with a batch size of 200. The Adam optimizer is used to optimize the categorical cross entropy function.

5.1. Optimization Parameter Reduction

As a baseline, the L_1-norm regularized three-level grid configuration requires the optimization of 128,625 B-spline basis coefficients in grid Level 1; 902,289 in Level 2; and 6,744,273 in Level 3—totaling in 7,775,187 optimization parameters. This optimization burden is greatly reduced by CNN control point classification. Because all test images result a different number control points in the "free" class, average control points per layer are reported. The average number of B-spline basis coefficients requiring optimization is reduced to 15,173 for grid Level 1; 60,081 for Level 2; and 355,390 for Level 3—totaling in 430,644 coefficients on average. Averages for both datasets across all three grid levels are provided in Table 1.

	Dataset	Before CNN	After CNN	Reduction
L1	Set 01	128,625	17,246	86%
	Set 02	128,625	13,618	89%
L2	Set 01	902,289	59,506	93%
	Set 02	902,289	60,512	93%
L3	Set 01	6,744,273	397,404	94%
	Set 02	6,744,273	323,880	95%

Table 1: Average number of B-spline parameters requiring optimization with and without CNN control point classification.

5.2. Prediction Quality

Figure 7(b-d, h-j, n-p) shows control point class prediction maps for three different randomly selected patients. Taking the control point labels produced by the L_1-norm regularized registration as the ground truth and the "free" class label to be positive, pixels shown in gray are true positives, those shown in purple are false positives, and those in cyan are false negatives. As illustrated, although predictions at finer grid levels produce an increased number of

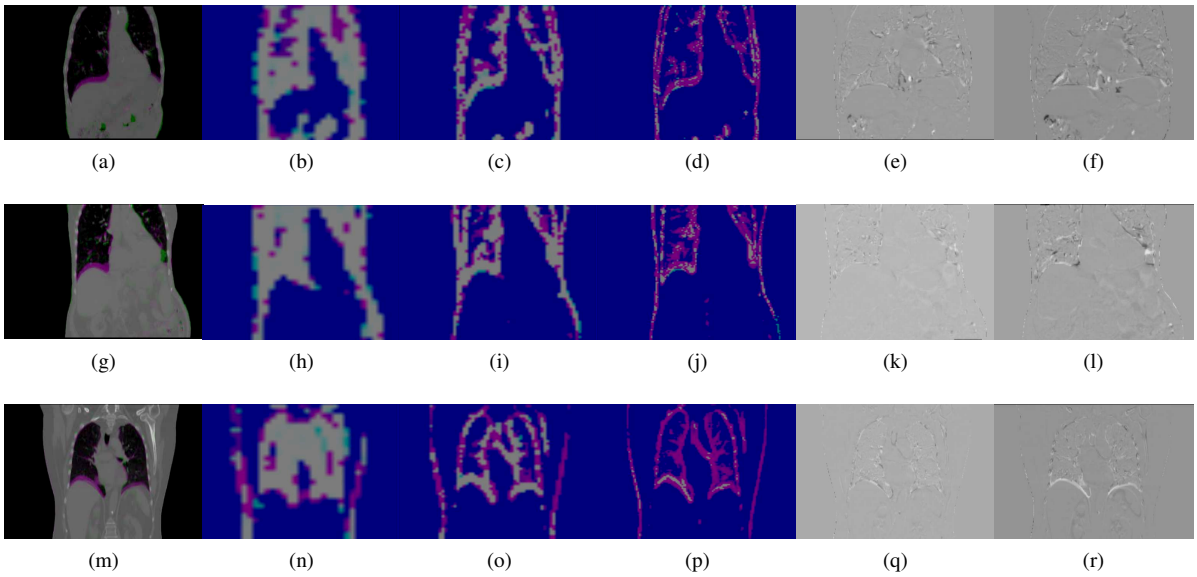


Figure 7: Illustration of CNN control point classification accuracy. (a) overlayed fixed and moving images prior to registration, (b) prediction map for grid Level 1 consisting of $35 \times 35 \times 35$ control points, (c) map for grid Level 2 with $67 \times 67 \times 67$ control points, and (d) map for grid Level 3 with $131 \times 131 \times 131$ control points. Prediction maps are gray when both the L_1-norm and CNN agree the coefficient is "free," cyan when the L_1-norm votes "free" and the CNN votes "locked," blue when both L1 norm and CNN agre that a coefficient is "locked," and purple when the L_1-norm votes "locked" and the CNN votes "free". (e) is the difference between the fixed image and the transformed moving image for the CNN driven registration process, and (f) is the same difference image for the L_1-norm driven process. (g-l) and (m-r) show the same data for two additional randomly selected datasets among those tested.

false positives, the majority of true positives are successfully classified by the CNN.

avg	Dataset	SN	SP	AC	AUC
L1	Set 01	0.91	0.94	0.94	0.94
	Set 02	0.94	0.97	0.97	0.96
L2	Set 01	0.91	0.97	0.97	0.95
	Set 02	0.92	0.97	0.97	0.95
L3	Set 01	0.89	0.95	0.95	0.93
	Set 02	0.90	0.96	0.96	0.94

Table 2: CNN classification performance quantified in terms of sensitivity (SN), specificity (SP), accuracy (AC), and area under curve (AUC) for each of the three grid levels tested.

Using the total number of true positive, true negative, false positive, and false negative control point classification instances, the classification performance of the CNN is expressed within Table 2 in terms of sensitivity:

$$SN = \frac{TP}{TP + FN} \tag{6}$$

specificity:

$$SP = \frac{TN}{FP + TN} \tag{7}$$

and accuracy:

$$AC = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}$$

The average sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC) are shown in Table 2 for each of the three grid levels employed in our thoracic 4D CT validation study. Prediction results are generally above 90% with an average accuracy of 95% or better for each grid level. Sources of error may include defects in the ground truth data. For example, ground truth deformation fields may potentially be improved on a case by case basis by extending the stopping condition beyond 20 iterations or by fine tuning the cost function parameters λ_R and λ_S.

5.3. Registration Results

Due to the decreased number of registration parameters requiring optimization under the proposed CNN driven algorithm coupled with the expressiveness of "free" class parameters, the SSD similarity metric decreased more rapidly

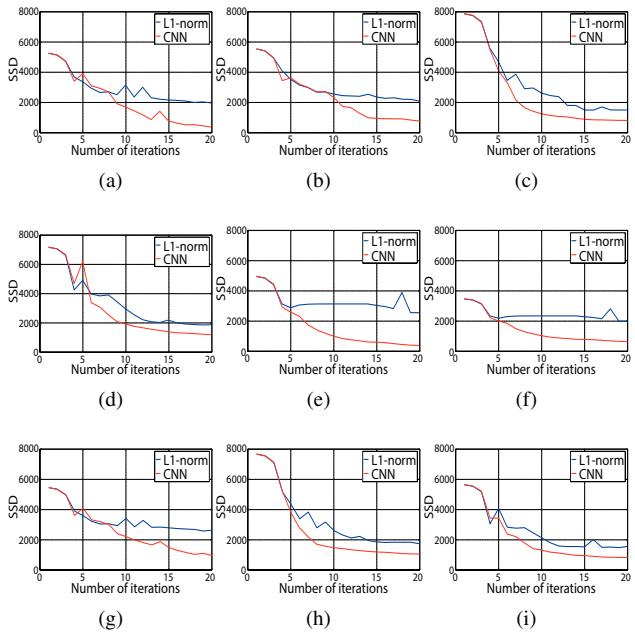


Figure 8: Comparison of registration quality between L_1-norm regularized multi-grid registration and the proposed CNN-based method. Results shown for nine randomly selected thoracic registrations: (a-f) full exhale to full inhale and (g-i) full inhale to full exhale.

in a fewer number of iterations when compared to the L_1-norm regularized baseline method. Nine randomly selected thoracic image registration results are shown in Figure 8 to illustrate the change in the SSD similarly metric with each L-BFGS-B optimization iteration. Each registration is performed using both L_1-norm regularized registration (shown in blue) and the proposed CNN driven registration (shown in red). The convergence rate of the proposed CNN method is faster for all nine cases. Additionally, higher quality registrations are achieved within these fewer iterations as shown in Figure 7(e, k, and q). In two cases, Figure 8(e and f), the L_1-norm regularized registration became prematurely trapped in local minima, whereas the CNN-based sparse B-spline managed to continue refining the quality of the deformation field.

6. Conclusion

In this paper we introduce a convolutional neural network driven multi-grid B-spline registration method. CNNs are used to to construct multi-level B-spline grids with predetermined sparsity prior to registration. This architecture not only addresses issues associated with the sequential, independent optimization of increasingly fine grid levels as performed by traditional hierarchical B-spline based methods, but also exhibits improved optimization efficiency and avoidance of local minima.

Acknowledgments: This material is based upon work supported by the National Science Foundation under Grant Numbers 1553436 (CAREER) and 1642380 (SI2-SSE).

References

[1] A. Andronache, M. von Siebenthal, G. Szekely, and P. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis (MIA)*, 12(1):3–15, 2008.

[2] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(6):567–585, 1989.

[3] C. Buerger, T. Schaeffter, and A. King. Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. *Medical Image Analysis (MIA)*, 15(4):551–564, 2011.

[4] R. Castillo, E. Castillo, R. Guerra, V. Johnson, T. McPhail, A. Garg, and T. Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology (PMB)*, 54(7):1849, 2009.

[5] S. Gu, X. Meng, F. Sciurba, H. Ma, J. Leader, N. Kaminski, D. Gur, and J. Pu. Bidirectional elastic image registration using B-spline affine transformation. *Computerized Medical Imaging and Graphics (CMIG)*, 38(4):306–314, 2014.

[6] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, 2002.

[7] P. Jiang and J. Shackleford. B-spline registration of neuroimaging modalites with map-reduce framework. In *International Conference on Brain Informatics and Health (BIH)*, pages 285–294. Springer, 2015.

[8] P. Jiang and J. Shackleford. An octree based approach to multi-grid B-spline registration. In *Proc. SPIE Medical Imaging*, volume 10133, pages 101330W–1, 2017.

[9] S. Klein, M. Staring, and J. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Transactions on Image Processing (TIP)*, 16(12):2879–2890, 2007.

[10] S. Lee, G. Wolberg, and S. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 3(3):228–244, 1997.

[11] H. Li, B. Manjunath, and S. K. Mitra. A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing (TIP)*, 4(3):320–334, 1995.

[12] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu. An artificial agent for robust image registration. In *Proc. Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, pages 4168–4175, 2017.

[13] S. Miao, Z. Wang, and R. Liao. A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging (TMI)*, 35(5):1352–1363, 2016.

[14] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin.

Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine (CMPB)*, 98(3):278–284, 2010.

[15] A. Pawar, Y. Zhang, Y. Jia, X. Wei, T. Rabczuk, C. Chan, and C. Anitescu. Adaptive FEM-based nonrigid image registration using truncated hierarchical B-splines. *Computers & Mathematics with Applications*, 72(8):2028–2040, 2016.

[16] D. Perperidis, R. H. Mohiaddin, and D. Rueckert. Spatio-temporal free-form registration of cardiac MR image sequences. *Medical Image Analysis (MIA)*, 9(5):441–456, 2005.

[17] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging (TMI)*, 22(6):730–741, 2003.

[18] D. Rueckert, I. Sonoda, C. Hayes, L. Hill, O. Leach, and J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging (TMI)*, 18(8):712–721, 1999.

[19] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 573–581. Springer, 2001.

[20] J. Shackleford, N. Kandasamy, and G. Sharp. On developing B-spline registration algorithms for multi-core processors. *Physics in Medicine and Biology (PMB)*, 55(21):6329, 2010.

[21] J. Shackleford, Q. Yang, A. Lourenco, N. Shusharina, N. Kandasamy, and G. Sharp. Analytic regularization of uniform cubic B-spline deformation fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 122–129. Springer, 2012.

[22] W. Shi, M. Jantsch, P. Aljabar, L. Pizarro, W. Bai, H. Wang, D. O'Regan, X. Zhuang, and D. Rueckert. Temporal sparse free-form deformations. *Medical Image Analysis (MIA)*, 17(7):779–789, 2013.

[23] W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K. Tung, P. Edwards, and D. Rueckert. Registration using sparse free-form deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 659–666. Springer, 2012.

[24] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing (TIP)*, 9(12):2083–2099, 2000.

[25] N. Tustison, B. Avants, and J. Gee. Directly manipulated free-form deformation image registration. *IEEE Transactions on Image Processing (TIP)*, 18(3):624–635, 2009.

[26] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.

[27] Z. Xie and G. E. Farin. Image registration using hierarchical B-splines. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 10(1):85–94, 2004.

[28] Y. Yin, E. Hoffman, and C. Lin. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Medical Physics*, 36(9):4213–4222, 2009.

[29] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis (MIA)*, 10(5):764–785, 2006.

[30] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Transactions on Medical Imaging (TMI)*, 30(10):1819–1828, 2011.