Evaluation of the DIRBoost algorithm on the EMPIRE10 dataset

Sascha E.A. Muenzing[1], Bram van Ginneken[2], Max A. Viergever[1], and Josien P.W. Pluim[1]

[1] Image Sciences Institute, University Medical Center Utrecht, The Netherlands,
sascha@isi.uu.nl,josien@isi.uu.nl
[2] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
b.vanginneken@rad.umcn.nl

Abstract. The DIRBoost algorithm is inspired by the theory on hypothesis boosting, well-known in the field of machine learning. Image registrations obtained by a registration algorithm are boosted by iteratively focusing the registration algorithm on erroneously registered regions. In this technical report we describe the application of the DIRBoost algorithm to the EMPIRE10 dataset employing three different registration algorithms: ANTS gSyN, NiftyReg, and DROP.

Keywords: Deformable image registration, hypothesis boosting, machine learning, pattern recognition

1 Introduction

The EMPIRE10 study has evaluated state-of-the-art registration methods for pulmonary intra-patient CT registration. Although many of those registration methods achieve excellent results on average, usually there are few scan pairs or distinct areas within one image where the registration fails. In this report we describe the application of the DIRBoost algorithm with three different registration algorithms previously evaluated on the EMPIRE10 dataset.

Deformable Image Registration. Registration of a moving image $I_M(\mathbf{x}) : \Omega_M \subset \mathbb{R}^D \mapsto \mathbb{R}$ to a fixed image $I_F(\mathbf{x}) : \Omega_F \subset \mathbb{R}^D \mapsto \mathbb{R}$, both of dimension D, is the problem of finding a displacement $\mathbf{u}(\mathbf{x})$ that makes $I_M(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ spatially aligned to $I_F(\mathbf{x})$. We define the obtained transformation field $\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$. The quality of alignment is defined by a distance or similarity measure, such as the normalized mutual information. If the underlying transformation model allows local deformations, i.e. nonlinear fields $\mathbf{T}(\mathbf{x})$, then we call it deformable image registration (DIR).

Boosting. The underlying idea of boosting is to combine simple "rules" to form an ensemble such that the performance of the single ensemble member is improved, i.e. "boosted". Let $h_1, h_2, \ldots, h_N$ be a set of hypotheses, and consider the composite ensemble hypothesis $f(\mathbf{x}) = \sum_{n=1}^{N} \alpha_n h_n(\mathbf{x})$. Here α_n denotes the coefficient with which the ensemble member h_n is combined; α_n and the hypothesis h_n are to be learned within the boosting procedure [8]. While boosting is not algorithmically constrained, most boosting algorithms consist of an iterative procedure where a hypothesis is learned with respect to (adaptively [4]) changed data distributions.

We hypothesize that the boosting theory can be transferred into the domain of deformable image registration. In this paper we present an iterative and adaptive algorithm for boosting deformable image registration (DIRBoost).

2 Methods

Details of the DIRBoost algorithm are described in [10]. For that reason, the following section only gives a brief description of the basic principles of DIRBoost.

In classifier boosting each ensemble member is weighted by a coefficient that is usually related to the accuracy of the single hypothesis. In adaptive boosting this means the weight of previously correctly classified instances decreases and the weight of misclassified instances increases. The aim of this adaptive weighting is to focus the training of subsequent classifiers on these difficult, yet incorrectly predicted instances. We propose a binary weighting scheme to achieve a similar effect for boosting of deformable image registration, and we define a weight field such that each voxel in the region of interest (lung segmentation) is assigned the binary weight one or zero, i.e. was previously misaligned or correctly registered. In practice that yields simple binary masks which can be applied to the input image pairs in order to focus the registration on previously incorrectly mapped voxels. That way, each boosting registration aims to improve the previous registration by replacing erroneous regions with displacements from the boosted displacement field. In order to boost an image registration algorithm, we require an estimate of local registration accuracy. For this purpose we employ an extended version of the method for automatic detection of registration errors described in [11].

3 Experiments

We demonstrate the DIRBoost algorithm on three different registration methods – ANTS, NiftyReg, DROP. NiftyReg and DROP are chosen because both are very fast DIR algorithms, which makes them excellent candidates for boosting because additional registrations can be performed relatively quickly. The ANTS gSyn method is computationally more expensive but has shown superior registration accuracy in several studies [7,12]. We consider the ANTS gSyn method as benchmark registration method. All three registration methods are of generic nature which allows application to a broad range of image data. Moreover, optimized configurations had been established and evaluated for intra-patient pulmonary lung CT registration [13]. We employ these registration configurations to obtain optimized base registrations.

3.1 ANTS

ANTS [2] is an open source software package (www.picsl.upenn.edu/ANTS) built on the ITK (Insight Segmentation and Registration Toolkit) framework [14]. ANTS comprises a suite of tools for image normalization and template building, in particular a diffeomorphic registration method with symmetric normalization [1].

Base registration The computation of the ANTS base registration is based on the registration set-up and parametrization proposed in [15]. An affine registration of the binary lung masks is used to initialize the deformable registration, which uses masked images such that the intensity values outside the lung masks were set to zero, and the intensity values inside the lung masks are normalized to $[0, 1]$ by linear intensity adjustment. Further, a symmetric variant of diffeomorphic transformation is employed, which ensures inverse consistency of the obtained registration [1]. A local cross correlation metric is calculated in a neighborhood around each voxel to accommodate the inhomogeneity of the density changes throughout the whole lung, and integrated over the lung volume as the overall similarity in the diffeomorphic transformation. For both the affine and diffeomorphic registration, the gradient descent is used in the optimization. Convergence during the optimization is achieved when the slope of linear regression of the cost values of last 12 iterations is close to zero. A multi-resolution approach is applied in both steps to accelerate computation speed and avoid trapped in local minimum.

We adapted the proposed multi-resolution setting [15] to further reduce computation time, conducting diffeomorphic registrations only until half the original image resolution, which setting we denote in the following by the acronym *ANTS2*.

DIR set-up for boosting We use the *ANTS2* parametrization for the deformable registration method as employed in the base registration, however, without preceding affine registration. ANTS (Release 1.5) does not support initialization of registrations with a transformation field, we therefore implement the indirect initialization approach described in Section 3.4.

3.2 NiftyReg

The NiftyReg registration package (sourceforge.net/projects/NiftyReg) contains a global and local registration algorithm. The global registration is based on a block-matching technique and the local registration is based on a B-Spline deformation model. B-Spline control point positions are optimized using a conjugate gradient ascent optimizer. The objective function is composed of normalized mutual information as a metric and optionally, the bending energy and the squared Jacobian determinant as penalty terms. The optimizer can be run until convergence of the registration parameters or until a maximum number of iterations is reached. NiftyReg includes a folding correction scheme that can be conducted concurrently with the registration process or separately as a post-registration process.

Base registration We use the registration set-up and parametrization proposed in [9] but perform all registrations until convergence. The proposed configuration makes use of lung masks to focus the registration on the lungs. The base registration consists of four stages: one global registration stage and three consecutive local registration stages. The result of the global registration is used to initialize control point positions of the first local stage. The aim of the first local registration stage is to quickly register the main structures in the lung. The second local stage aims at aligning the border of the lungs. And the goal of the final stage is the detailed alignment of the entire lung. The base registration is a multi-resolution approach consisting of four stages: one global registration stage and three consecutive local registration stages.

DIR set-up for boosting We use the same parametrization as employed in the base registration but require only a reduced set-up consisting of the two local registration stages because DIRBoost is initialized by the transformation field of the base registration. The folding correction is equal to the folding correction in stage three of the base registration. In addition we require auxiliary external computations to transform the deformation field into a coarse B-Spline representation with which we initialize the DIR algorithm.

3.3 DROP

DROP (mrf-registration.net) is a software package for deformable image registration using discrete optimization. The registration problem is formulated using a discrete Markov random field objective function [6]. It is based on the assumption that a dense deformation field can be expressed using a small number of control points (registration grid) and an interpolation strategy. The registration cost is then expressed using a discrete sum over image costs (using an arbitrary similarity measure) projected on the control points, and a smoothness term that penalizes local deviations on the deformation field according to a neighborhood system on the grid. The search space is quantized resulting in a fully discrete model. Efficient linear programming using the primal dual principles is then employed to recover the lowest potential of the cost function. In order to account for large deformations and produce results on a high-resolution level, a multi-scale incremental approach is considered where the optimal solution is iteratively updated.

Base registration The computation of the DROP base registration is based on the registration set-up and parametrization proposed in [5], however we adapted the settings to further optimize DROP for the particular breath-hold scan pairs. We employ the lung segmentations and use masked scans to focus the registration on the interior of the lung. Further, we use normalized cross correlation as similarity metric with the weighting parameter α set to zero along with the compositional update rule in order to obtain bijective transformations. The maximum number of iterations was never reached, that is we performed each registration of the four grid levels until convergence.

DIR set-up for boosting We use the same parametrization employed in the base registration but require only a maximum of ten iterations per grid level. We determined in preceding evaluations of the base registration that unlimited iterations yield no significant improvement in registration accuracy compared with ten iterations. DROP (version 1.6) does not support initialization of registrations with a transformation field, we therefore implement the indirect initialization approach described in Section 3.4.

3.4 Indirect registration initialization

The proposed DIRBoost method iteratively improves registrations by boosting the displacement field obtained in the previous iteration. In cases where the employed DIR algorithm does not provide a registration initialization with a displacement field, the following approach for indirect initialization is implemented. First, the obtained displacement field $\mathbf{u}_n$ is smoothed with a Gaussian kernel with σ equal to the coarsest voxel spacing applied in the registration pipeline. Second, the moving image I_M is warped using the smoothed displacement field $\mathbf{u}_\sigma$ yielding the registered image $I_{R\sigma}$. Third, image registration is performed employing the image pair $< I_F,\ I_{R\sigma} >$ instead of $< I_F,\ I_M >$. Fourth, the obtained displacement field $\mathbf{u}_{n+1}$ is composed with $\mathbf{u}_\sigma$ to account for the initial warp of I_M. In order to boost registrations optimally, the DIR algorithm should have access to the original image information of I_M, which might be distorted in I_R beyond recognition and therefore might prevent the DIR algorithm from recovering the underlying deformation between I_F and I_M. The use of $I_{R\sigma}$ instead of I_R as surrogate initialization is crucial to alleviate possible image distortions.

3.5 Registration error detection

We employ an automatic method for quality assessment of medical image registration [11] in order to locate registration errors. The method is based on supervised learning of local alignment patterns, which are captured by statistical image features at automatically detected landmarks.

Supervised quality assessment For supervised learning a training database S is established. It combines information from three datasets: a) reference landmark matchings, b) reference image registrations, and c) statistical image features. The set of reference landmark matchings consists of landmarks L_F on the fixed image I_F and their corresponding location L_M in the moving image I_M. The set of reference image registrations consists of transformation fields $\mathbf{T}$ obtained by the particular registration method on the particular image data, on which registration errors shall be detected. The set of statistical image features (FS) contains for each landmark of L_F the corresponding feature values that are computed based on different image feature types. Gaussian, correlation, and entropy features are calculated from the intensity images (I_F and I_R), and deformation features are computed on the transformation fields $\mathbf{T}$.

Based on the database S, a classifier cascade is trained to classify local alignment patterns into three quality categories: correct (CA), poor (PA), and wrong alignment (WA). The quality categories are based on the landmark registration error (LRE), i.e. the amount of misalignment between a registered landmark position $\mathbf{T}(L_F)$ and the corresponding reference matching L_M. To automatically classify a previously unseen registration, first, a large set of landmarks L_U is automatically detected on corresponding I_F. Second, image features are extracted according to FS for each landmark of L_U. Third, the trained classifier is applied to each landmark sample yielding a quality assignment for each landmark. Fourth, the landmark-based quality estimate is interpolated to obtain a dense quality estimate. Given a sufficiently large number of well-distributed landmarks L_U, the initially sparse quality estimate can be converted into reliable voxel-based predictions. We opt for a Voronoi decomposition for spatial interpolation of landmark-based quality estimates. In this manner, a dense assessment of local image registration quality is obtained where every voxel of the region-of-interest is assigned a quality category.

Image features were extracted on ROI masked images, similarly to [11]. We adapted scales and added features to better capture alignment patterns of registrations from inspiration-expiration scans. All cubic subvolumes are sampled in the range from 5 to 100mm. The set of image features described in [11] has been extended by the following geometrical, deformation-based and cubic subvolume intensity features. Normalized landmark position $NLP(\mathbf{x})$: physical position normalized by bounding box of ROI in M_F. Distance of landmark to ROI boundary $DTB(\mathbf{x})$: computed based on distance transformation of M_F and M_M.

Normalized Jacobian map $NJM(\mathbf{x};\sigma)$: $|J\left(\mathbf{u}(\mathbf{x}) * G(\mathbf{x};\sigma)\right)|/vr$ with vr the volume ratio of the ROI between I_F and I_M. Mean of absolute and squared intensity difference $MID(\mathbf{x};vol)$, $MSID(\mathbf{x};vol)$. In addition, texture features $GLCM(\mathbf{x};\sigma)$ based on gray-level co-occurrence matrices (GLCM) are included for the application on the mixed image data of the EMPIRE10 dataset, in order to distinguish different image categories based on contrast, dissimilarity and homogeneity features.

In the following we refer by RED to this entity of automatic registration error detection.

3.6 Training & application of RED

To properly train a RED classifier for the mixed image data of the EMPIRE10 set, we composed a mixed reference standard set consisting of three different reference datasets: 5 NELSON follow-up scan pairs [11], 5 NELSON breathhold scan pairs [10], and 10 COPDgene breathhold scan pairs [3].

We computed reference image registrations for each reference image data and for each registration method. Then, for each registration method, a specific learning dataset S was established. We set the quality categories (CA,PA,WA), based on the landmark registration error (LRE), to the following values: CA $\doteq$ $\{\text{LRE} \leq 2 \text{ mm}\}$, PA $\doteq$ $\{2 < \text{LRE} < 5 \text{ mm}\}$, WA $\doteq$ $\{\text{LRE} \geq 5 \text{ mm}\}$. However, different values could be chosen to define registration quality and registration errors. The random forests (RF) classifier was employed, using 1000 trees and the default feature subsampling setting. We limited the size of the feature subsets so that the performance of the optimal feature-model is equally well or better than employing the whole set of 190 features. On average, about 50 features were selected in both domains of the two stage classification cascade.

For application of the RED classifier, first, a set of statistical landmarks L_U is automatically detected on the fixed image I_F of scan pair k. Second, statistical image feature FS are extracted at landmarks L_U, and classified by the trained RED classifier. See [11] for further details on the supervised method for registration error detection.

3.7 Computation times

The proposed DIRBoost algorithm in the first place aims to improve registration accuracy. Minimization of computation times was not in the scope of this work, and the implementation has not been explicitly optimized for speed. On a PC with an Intel Core 2 Quad CPU (2.66GHz) and 8GB RAM, one iteration of DIRBoost took about 5 minutes, in addition to the time required to perform the deformable registration.

4 Results & Discussion

The ANTS2, NiftyReg, and DROP registrations that were deemed erroneous by the automatic registration error detection underwent boosting. In this manner, eight ANTS2, nine NiftyReg, and eleven DROP base registrations were subjected to boosting. Tables 1,3,5 show the evaluation results of the base registrations, and Tables 2,4,6 show the evaluation results after boosting. Table 7 lists the evaluation results of the ANTS gSyN benchmark registrations. The rankings and final placements are from a total of 34 competing algorithms (September, 19th, 2013).

The EMPIRE10 evaluation results show that DIRBoost could improve registrations of each of the DIR methods. Overall, the boosted registrations are superior to the base registrations, according to the EMPIRE10 ranking scheme. However, the registration results of ANTS and NiftyReg reported previously on EMPIRE10 are superior to the results obtained in our experiments, although we tried to emulate these results by choosing parameter settings and registration setups as described in [15,9]. This outcome demonstrates that the end result of the boosting will depend on the base registration method, i.e. the registration algorithm and the registration setup and parameter settings.

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.16	32.50	0.14	20.50	1.44	12.50	0.00	15.00
02	0.00	16.00	0.00	17.00	0.36	13.00	0.00	17.00
03	0.00	10.00	0.00	15.00	0.28	2.00	0.00	17.50
04	0.00	5.00	0.00	19.00	0.83	7.00	0.00	17.00
05	0.00	16.50	0.00	17.00	0.00	7.00	0.00	17.00
06	0.00	18.00	0.00	6.50	0.28	11.00	0.00	18.50
07	0.00	7.50	0.16	5.50	1.14	5.50	0.00	14.00
08	0.01	31.00	0.00	9.00	0.61	9.00	0.00	17.00
09	0.00	3.50	0.00	8.50	0.50	4.00	0.00	16.50
10	0.00	3.50	0.00	17.50	1.98	20.00	0.00	32.00
11	0.03	26.50	0.01	10.50	0.70	13.50	0.00	16.00
12	0.00	33.00	0.00	29.00	0.00	5.00	0.00	18.50
13	0.00	5.50	0.06	7.00	0.75	10.00	0.00	16.50
14	0.00	6.50	1.00	5.50	1.80	13.50	0.00	23.50
15	0.00	11.50	0.00	7.50	0.62	11.00	0.00	17.00
16	0.00	6.50	0.00	4.00	0.91	10.00	0.00	18.00
17	0.00	12.00	0.06	33.00	0.60	5.00	0.00	17.50
18	0.00	6.50	0.03	3.50	1.10	7.50	0.00	26.50
19	0.00	18.50	0.00	12.50	0.45	11.00	0.00	17.00
20	0.00	25.50	0.30	4.50	1.02	5.50	0.00	14.50
21	0.10	28.50	0.25	6.50	1.27	4.50	0.00	14.00
22	0.00	18.00	0.00	7.50	0.74	6.00	0.00	16.50
23	0.00	12.50	0.03	9.50	0.71	15.00	0.00	17.50
24	0.00	8.00	0.00	14.50	0.74	7.00	0.00	17.00
25	0.00	16.50	0.00	16.00	0.01	8.00	0.00	16.00
26	0.00	12.50	0.00	12.50	0.34	9.00	0.00	18.50
27	0.00	10.00	0.00	5.50	0.37	7.00	0.00	17.50
28	0.02	29.50	1.57	10.50	1.55	10.50	0.00	15.50
29	0.00	8.50	0.00	16.00	1.19	21.00	0.00	16.50
30	0.00	32.00	0.00	15.50	0.00	5.00	0.00	17.50
Avg	0.01	15.71	0.12	12.21	0.74	9.20	0.00	17.76
Average Ranking Overall								13.72
Final Placement								6.5

Table 1. ANTS2 base registrations. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.05	27.00	0.00	4.00	1.46	14.00	0.00	15.00
02	0.00	16.00	0.00	17.00	0.36	13.00	0.00	17.00
03	0.00	10.00	0.00	15.00	0.28	2.00	0.00	17.50
04	0.00	5.00	0.00	19.00	0.83	7.00	0.00	17.00
05	0.00	16.50	0.00	17.00	0.00	7.00	0.00	17.00
06	0.00	18.00	0.00	6.50	0.28	11.00	0.00	18.50
07	0.00	6.00	0.29	8.00	1.06	4.00	0.00	14.00
08	0.01	31.00	0.00	9.00	0.61	9.00	0.00	17.00
09	0.00	3.50	0.00	8.50	0.50	4.00	0.00	16.50
10	0.00	3.50	0.00	17.50	1.98	20.00	0.00	32.00
11	0.03	25.00	0.00	6.00	0.62	7.00	0.00	16.00
12	0.00	33.00	0.00	29.00	0.00	5.00	0.00	18.50
13	0.00	5.50	0.06	7.00	0.75	10.00	0.00	16.50
14	0.00	5.00	0.78	4.00	0.99	4.00	0.00	11.00
15	0.00	11.50	0.00	7.50	0.62	11.00	0.00	17.00
16	0.00	6.50	0.00	4.00	0.91	10.00	0.00	18.00
17	0.00	12.00	0.06	33.00	0.60	5.00	0.00	17.50
18	0.00	2.00	0.03	2.00	1.04	4.00	0.00	13.00
19	0.00	18.50	0.00	12.50	0.45	11.00	0.00	17.00
20	0.00	22.00	0.19	2.00	1.04	8.00	0.00	14.50
21	0.06	25.00	0.18	4.00	1.33	6.00	0.00	14.00
22	0.00	18.00	0.00	7.50	0.74	6.00	0.00	16.50
23	0.00	12.50	0.03	9.50	0.71	15.00	0.00	17.50
24	0.00	8.00	0.00	14.50	0.74	7.00	0.00	17.00
25	0.00	16.50	0.00	16.00	0.01	8.00	0.00	16.00
26	0.00	12.50	0.00	12.50	0.34	9.00	0.00	18.50
27	0.00	10.00	0.00	5.50	0.37	7.00	0.00	17.50
28	0.00	27.00	0.35	1.00	1.27	2.00	0.00	15.50
29	0.00	8.50	0.00	16.00	1.19	21.00	0.00	16.50
30	0.00	32.00	0.00	15.50	0.00	5.00	0.00	17.50
Avg	0.00	14.91	0.06	11.01	0.70	8.39	0.00	16.90
Average Ranking Overall								12.80
Final Placement								4

Table 2. ANTS2 boosted. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms. Base registrations subjected to boosting: 1,7,11,14,18,20,21,28.

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.00	4.50	0.01	13.00	1.32	9.00	0.00	15.00
02	0.00	16.00	0.00	17.00	0.42	19.50	0.00	17.00
03	0.00	10.00	0.00	15.00	0.43	22.50	0.00	17.50
04	0.00	17.50	0.00	19.00	1.84	26.50	0.00	17.00
05	0.00	16.50	0.00	17.00	0.08	24.50	0.00	17.00
06	0.00	18.00	0.00	32.50	0.28	8.50	0.00	18.50
07	0.00	2.50	1.88	29.00	1.40	13.00	0.00	14.00
08	0.00	7.00	0.01	17.00	0.76	18.00	0.00	17.00
09	0.00	3.50	0.05	33.50	0.56	20.50	0.00	16.50
10	0.00	20.00	0.00	17.50	2.19	22.00	0.00	15.50
11	0.00	8.00	0.02	18.00	0.97	19.00	0.00	16.00
12	0.00	26.50	0.00	13.00	0.05	18.50	0.00	18.50
13	0.00	13.50	0.08	19.50	0.64	1.50	0.00	16.50
14	0.53	38.00	4.00	25.00	1.43	10.00	0.00	11.00
15	0.00	11.50	0.03	31.50	0.65	22.50	0.00	17.00
16	0.00	18.50	0.00	12.00	1.00	16.50	0.00	18.00
17	0.00	12.00	0.04	13.50	0.59	1.50	0.00	17.50
18	0.00	3.00	0.50	15.00	1.23	11.00	0.00	13.00
19	0.00	18.50	0.00	28.00	0.47	19.50	0.00	17.00
20	0.00	6.50	1.06	14.50	1.07	12.50	0.00	14.50
21	0.00	4.00	0.84	20.00	1.59	11.00	0.00	14.00
22	0.00	3.00	0.03	25.50	0.81	12.50	0.00	16.50
23	0.00	12.50	0.10	22.00	0.63	6.50	0.00	17.50
24	0.00	22.50	0.00	14.50	1.06	20.50	0.00	17.00
25	0.00	16.50	0.00	16.00	0.14	21.50	0.00	16.00
26	0.00	12.50	0.00	12.50	0.45	19.50	0.00	18.50
27	0.00	10.00	0.00	24.50	0.39	11.50	0.00	17.50
28	0.00	7.00	2.69	18.00	3.17	26.00	0.00	15.50
29	0.00	8.50	0.00	16.00	1.15	18.50	0.00	16.50
30	0.00	15.50	0.00	15.50	0.09	16.50	0.00	17.50
Avg	0.01	12.78	0.38	19.50	0.90	16.01	0.00	16.35
Average Ranking Overall								16.16
Final Placement								15

Table 3. NiftyReg base registrations. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.00	4.50	0.01	11.00	1.19	4.00	0.00	15.00
02	0.00	16.00	0.00	17.00	0.42	19.50	0.00	17.00
03	0.00	10.00	0.00	15.00	0.43	22.50	0.00	17.50
04	0.00	17.50	0.00	19.00	1.84	26.50	0.00	17.00
05	0.00	16.50	0.00	17.00	0.08	24.50	0.00	17.00
06	0.00	18.00	0.00	32.50	0.28	8.50	0.00	18.50
07	0.08	32.00	1.49	27.00	1.24	9.00	0.00	14.00
08	0.00	7.00	0.03	21.00	0.53	4.00	0.00	17.00
09	0.00	3.50	0.05	33.50	0.56	20.50	0.00	16.50
10	0.00	18.00	0.00	17.50	1.72	18.00	0.00	15.50
11	0.00	16.00	0.01	17.00	0.82	16.00	0.00	16.00
12	0.00	26.50	0.00	13.00	0.05	18.50	0.00	18.50
13	0.00	13.50	0.08	19.50	0.64	1.50	0.00	16.50
14	0.03	26.00	3.61	24.00	1.28	9.00	0.00	11.00
15	0.00	11.50	0.03	31.50	0.65	22.50	0.00	17.00
16	0.00	18.50	0.00	12.00	1.00	16.50	0.00	18.00
17	0.00	12.00	0.04	13.50	0.59	1.50	0.00	17.50
18	0.00	13.00	0.25	13.00	1.07	5.00	0.00	13.00
19	0.00	18.50	0.00	28.00	0.47	19.50	0.00	17.00
20	0.00	6.50	1.06	14.50	1.07	12.50	0.00	14.50
21	0.00	6.00	0.46	14.00	1.37	7.00	0.00	14.00
22	0.00	3.00	0.03	25.50	0.81	12.50	0.00	16.50
23	0.00	12.50	0.10	22.00	0.63	6.50	0.00	17.50
24	0.00	22.50	0.00	14.50	1.06	20.50	0.00	17.00
25	0.00	16.50	0.00	16.00	0.14	21.50	0.00	16.00
26	0.00	12.50	0.00	12.50	0.45	19.50	0.00	18.50
27	0.00	10.00	0.00	24.50	0.39	11.50	0.00	17.50
28	0.00	3.50	1.40	8.00	1.43	7.00	0.00	15.50
29	0.00	8.50	0.00	16.00	1.15	18.50	0.00	16.50
30	0.00	15.50	0.00	15.50	0.09	16.50	0.00	17.50
Avg	0.00	13.85	0.29	18.83	0.78	14.01	0.00	16.35
Average Ranking Overall								15.76
Final Placement								12

Table 4. NiftyReg boosted. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms. Base registrations subjected to boosting: 1,7,8,10,11,14,18,21,28

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.00	15.00	0.49	26.00	3.39	28.00	0.00	15.00
02	0.00	16.00	0.00	17.00	0.46	23.50	0.00	17.00
03	0.00	23.50	0.00	32.50	0.53	25.50	0.00	17.50
04	0.00	19.50	0.00	19.00	1.30	20.50	0.00	17.00
05	0.00	16.50	0.00	34.50	0.06	22.50	0.00	17.00
06	0.00	18.00	0.00	26.50	0.38	26.50	0.00	18.50
07	0.00	16.00	1.13	23.00	4.39	31.00	0.00	14.00
08	0.00	7.00	0.80	30.00	1.40	28.00	0.00	17.00
09	0.00	22.50	0.02	29.50	0.67	27.50	0.00	16.50
10	0.00	23.50	0.00	17.50	3.65	29.50	0.00	15.50
11	0.00	13.00	0.90	31.00	1.91	31.00	0.00	16.00
12	0.00	13.00	0.00	31.50	0.43	27.50	0.00	18.50
13	0.00	17.00	0.15	32.00	1.09	25.00	0.00	16.50
14	0.00	15.00	4.90	28.00	5.38	30.00	0.00	11.00
15	0.00	11.50	0.00	27.50	0.66	24.50	0.00	17.00
16	0.00	6.50	0.02	17.50	1.10	20.50	0.00	18.00
17	0.00	12.00	0.06	29.00	1.04	24.50	0.00	17.50
18	0.00	17.00	3.93	29.00	4.25	31.00	0.00	13.00
19	0.00	18.50	0.00	28.00	0.62	26.50	0.00	17.00
20	0.00	17.00	7.73	35.00	5.84	30.00	0.00	14.50
21	0.00	12.00	0.94	23.00	3.36	21.00	0.00	14.00
22	0.00	3.00	0.04	27.50	1.28	29.50	0.00	16.50
23	0.00	12.50	0.06	17.50	0.92	23.50	0.00	17.50
24	0.00	19.50	0.00	31.50	1.31	28.50	0.00	17.00
25	0.00	16.50	0.00	32.50	0.21	25.50	0.00	16.00
26	0.00	12.50	0.00	4.50	0.58	25.50	0.00	18.50
27	0.00	10.00	0.01	28.50	0.58	23.50	0.00	17.50
28	0.00	16.00	2.32	15.00	3.71	27.00	0.00	15.50
29	0.00	8.50	0.00	16.00	1.54	24.50	0.00	16.50
30	0.00	15.50	0.00	15.50	0.32	23.50	0.00	17.50
Avg	0.00	14.80	0.78	25.18	1.74	26.16	0.00	16.35
Average Ranking Overall								20.62
Final Placement								25

Table 5. DROP base registrations. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.00	9.00	0.10	19.00	2.33	20.00	0.00	15.00
02	0.00	16.00	0.00	17.00	0.46	23.50	0.00	17.00
03	0.00	23.50	0.00	32.50	0.53	25.50	0.00	17.50
04	0.00	19.50	0.00	19.00	1.30	20.50	0.00	17.00
05	0.00	16.50	0.00	34.50	0.06	22.50	0.00	17.00
06	0.00	18.00	0.00	26.50	0.38	26.50	0.00	18.50
07	0.00	5.00	0.57	13.00	2.42	22.00	0.00	14.00
08	0.00	7.00	0.24	25.00	1.29	26.00	0.00	17.00
09	0.00	22.50	0.02	29.50	0.67	27.50	0.00	16.50
10	0.00	23.50	0.00	17.50	3.65	29.50	0.00	15.50
11	0.00	12.00	0.85	30.00	1.31	25.00	0.00	16.00
12	0.00	13.00	0.00	31.50	0.43	27.50	0.00	18.50
13	0.00	12.00	0.15	31.00	1.07	23.00	0.00	16.50
14	0.00	13.00	4.10	26.00	2.74	22.00	0.00	11.00
15	0.00	11.50	0.00	27.50	0.66	24.50	0.00	17.00
16	0.00	6.50	0.02	17.50	1.10	20.50	0.00	18.00
17	0.00	12.00	0.06	29.00	1.04	24.50	0.00	17.50
18	0.00	22.00	1.70	22.00	2.25	19.00	0.00	13.00
19	0.00	18.50	0.00	28.00	0.62	26.50	0.00	17.00
20	0.00	18.00	1.90	19.00	2.24	23.00	0.00	14.50
21	0.00	14.00	0.60	17.00	2.59	18.00	0.00	14.00
22	0.00	3.00	0.04	27.50	1.28	29.50	0.00	16.50
23	0.00	12.50	0.06	17.50	0.92	23.50	0.00	17.50
24	0.00	19.50	0.00	31.50	1.31	28.50	0.00	17.00
25	0.00	16.50	0.00	32.50	0.21	25.50	0.00	16.00
26	0.00	12.50	0.00	4.50	0.58	25.50	0.00	18.50
27	0.00	10.00	0.01	28.50	0.58	23.50	0.00	17.50
28	0.00	17.00	1.15	6.00	2.27	18.00	0.00	15.50
29	0.00	8.50	0.00	16.00	1.54	24.50	0.00	16.50
30	0.00	15.50	0.00	15.50	0.32	23.50	0.00	17.50
Avg	0.00	14.26	0.38	23.05	1.27	23.96	0.00	16.35
Average Ranking Overall								19.40
Final Placement								21

Table 6. DROP boosted. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms. Base registrations subjected to boosting: 1,7,8,11,13,14,18,20,21,22,28.

Scan Pair	Lung Boundaries Score	Lung Boundaries Rank	Fissures Score	Fissures Rank	Landmarks Score	Landmarks Rank	Singularities Score	Singularities Rank
01	0.16	32.50	0.14	20.50	1.44	12.50	0.00	15.00
02	0.00	16.00	0.00	17.00	0.36	13.00	0.00	17.00
03	0.00	10.00	0.00	15.00	0.28	2.00	0.00	17.50
04	0.00	5.00	0.00	19.00	0.83	7.00	0.00	17.00
05	0.00	16.50	0.00	17.00	0.00	7.00	0.00	17.00
06	0.00	18.00	0.00	6.50	0.28	11.00	0.00	18.50
07	0.00	7.50	0.16	5.50	1.14	5.50	0.00	14.00
08	0.01	31.00	0.00	9.00	0.61	9.00	0.00	17.00
09	0.00	3.50	0.00	8.50	0.50	4.00	0.00	16.50
10	0.00	3.50	0.00	17.50	1.98	20.00	0.00	32.00
11	0.03	26.50	0.01	10.50	0.70	13.50	0.00	16.00
12	0.00	33.00	0.00	29.00	0.00	5.00	0.00	18.50
13	0.00	5.50	0.06	7.00	0.75	10.00	0.00	16.50
14	0.00	6.50	1.00	5.50	1.80	13.50	0.00	23.50
15	0.00	11.50	0.00	7.50	0.62	11.00	0.00	17.00
16	0.00	6.50	0.00	4.00	0.91	10.00	0.00	18.00
17	0.00	12.00	0.06	33.00	0.60	5.00	0.00	17.50
18	0.00	6.50	0.03	3.50	1.10	7.50	0.00	26.50
19	0.00	18.50	0.00	12.50	0.45	11.00	0.00	17.00
20	0.00	25.50	0.30	4.50	1.02	5.50	0.00	14.50
21	0.10	28.50	0.25	6.50	1.27	4.50	0.00	14.00
22	0.00	18.00	0.00	7.50	0.74	6.00	0.00	16.50
23	0.00	12.50	0.03	9.50	0.71	15.00	0.00	17.50
24	0.00	8.00	0.00	14.50	0.74	7.00	0.00	17.00
25	0.00	16.50	0.00	16.00	0.01	8.00	0.00	16.00
26	0.00	12.50	0.00	12.50	0.34	9.00	0.00	18.50
27	0.00	10.00	0.00	5.50	0.37	7.00	0.00	17.50
28	0.02	29.50	1.57	10.50	1.55	10.50	0.00	15.50
29	0.00	8.50	0.00	16.00	1.19	21.00	0.00	16.50
30	0.00	32.00	0.00	15.50	0.00	5.00	0.00	17.50
Avg	0.01	15.71	0.12	12.21	0.74	9.20	0.00	17.76
Average Ranking Overall								13.72
Final Placement								6.5

Table 7. ANTS benchmark registrations. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

References

1. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12(1), 26–41 (2008)
2. Avants, B., Tustison, N., Song, G., Gee, J.: ANTS, Advanced Normalization Tools (2013)
3. Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A.M., Ludwig, M.S., Guerrero, T.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Physics in Medicine and Biology 58(9), 2861 (2013)
4. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55, 119–139 (1997)
5. Glocker, B., Komodakis, N., Paragios, N., Navab, N.: Non-rigid Registration using Discrete MRFs: Application to Thoracic CT Images. In: Medical Image Analysis for the Clinic - A Grand Challenge 2010. pp. 147–154 (2010)
6. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
7. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46(3), 786–802 (2009)
8. Meir, R., Rätsch, G.: An introduction to boosting and leveraging. In: Advanced Lectures on Machine Learning. pp. 118–183. Springer-Verlag New York, Inc., New York, NY, USA (2003)
9. Modat, M., McClelland, J., Ourselin, S.: Lung registration using the NiftyReg package. In: Medical Image Analysis for the Clinic - A Grand Challenge 2010. pp. 33–42 (September 2010)
10. Muenzing, S.E.A., van Ginneken, B., Pluim, J.P.W.: DIRBoost: An algorithm for boosting deformable image registration. In: IEEE International Symposium on Biomedical Imaging. pp. 1339–1342 (2012)
11. Muenzing, S.E., van Ginneken, B., Murphy, K., Pluim, J.P.: Supervised quality assessment of medical image registration: Application to intra-patient CT lung registration. Medical Image Analysis 16(8), 1521 – 1531 (2012)
12. Murphy, K., van Ginneken, B., Klein, S., Staring, M., de Hoop, B., Viergever, M., Pluim, J.P.: Semi-automatic construction of reference standards for evaluation of image registration. Medical Image Analysis 15, 71–84 (2011)
13. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M., Schnabel, J., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J., Ourselin, S., Muenzing, S., Viergever, M., Nigris, D.D., Collins, D., Arbel, T., Peroni, M., Li, R., Sharp, G., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J., Staring, M., Klein, S., Stoel, B., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D., Pluim, J.: Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. IEEE Transactions on Medical Imaging 30, 1901–1920 (2011)

14. National Library of Medicine: ITK, Insight Segmentation and Registration Toolkit (2013)
15. Song, G., Tustison, N., Avants, B., Gee, J.C.: Lung CT image registration using diffeomorphic transformation models. In: Medical Image Analysis for the Clinic - A Grand Challenge 2010. pp. 23–32 (2010)