Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. Please note that you are not allowed to share this article on other platforms, but can link to it. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication or parts of it other than authorised under this licence or copyright law is prohibited. Neither Radboud University nor the authors of this publication are liable for any damage resulting from your (re)use of this publication.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.
DIRBoost–An algorithm for boosting deformable image registration: Application to lung CT intra-subject registration

Sascha E.A. Muenzing, Bram van Ginneken, Max A. Viergever, Josien P.W. Pluim

Abstract

We introduce a boosting algorithm to improve on existing methods for deformable image registration (DIR). The proposed DIRBoost algorithm is inspired by the theory on hypothesis boosting, well known in the field of machine learning. DIRBoost utilizes a method for automatic registration error detection to obtain estimates of local registration quality. All areas detected as erroneously registered are subjected to boosting, i.e. undergo iterative registrations by employing boosting masks on both the fixed and moving image. We validated the DIRBoost algorithm on three different DIR methods (ANTS gSyn, NiftyReg, and DROP) on three independent reference datasets of pulmonary image scan pairs. DIRBoost reduced registration errors significantly and consistently on all reference datasets for each DIR algorithm, yielding an improvement of the registration accuracy by 5–34% depending on the dataset and the registration algorithm employed.

1. Introduction

Accurate registration of medical images is key to medical image analysis. Many successful registration algorithms have been proposed but generally in a limited setting, and even then there are instances where the registration fails locally. In this paper we propose a boosting algorithm for deformable image registration (DIRBoost) which aims to boost these non-systematic registration failures. Based on estimates of local registration quality, DIRBoost focuses on the registration of erroneous regions in an iterative manner, similar to the well-established boosting of classifiers (Meir and Rätsch, 2003). This paper is based on an earlier publication (Muenzing et al., 2012b), in which we presented a proof of concept. DIRBoost is validated for three registration algorithms (ANTS gSyn, NiftyReg, DROP), and on three independent reference datasets of intra-patient pulmonary registrations (NELSON, COPDgene, EMPIRE10).

1.1. Deformable image registration

Registration of a moving image \( I_f(x) : \Omega_f \subset \mathbb{R}^D \rightarrow \mathbb{R} \) to a fixed image \( I_t(x) : \Omega_t \subset \mathbb{R}^D \rightarrow \mathbb{R} \), both of dimension \( D \), is the problem of finding a displacement field \( u(x) \) that makes \( I_f(x + u(x)) \) spatially aligned to \( I_t(x) \). We define the obtained transformation \( T(x) = x + u(x) \) and the registered image \( I_R(x) = I_t(x) \circ T(x) \). If the underlying transformation model allows local deformations, i.e. non-linear fields \( u(x) \), then we call it deformable image registration (DIR).

1.2. 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 \( f(x) = \sum_{i=1}^{N} \alpha_i h_i(x) \). Here \( \alpha_i \) denotes the coefficient with which the ensemble member \( h_i \) is weighted; \( \alpha \) and the hypothesis \( h \) are to be learned within the boosting procedure (Meir and Rätsch, 2003).

Dietterich (2002) gives the following explanation why ensemble methods work in the domain of machine learning. Algorithms that output only a single hypothesis suffer from three problems that can be partly overcome by ensemble methods: the statistical problem, the representational problem, and the computational problem. The statistical problem arises when the learning algorithm searches a space of hypotheses that is too large for the amount of available training data. In such cases, there may be several different hypotheses that all give the same accuracy on the training data, and the learning algorithm must choose one of these to output. There is a risk that the chosen hypothesis will not fit well on future data. An ensemble of classifiers can reduce this risk. The
representational problem arises when the hypothesis space does not contain any hypotheses that are good approximations to the true function. In some cases, an ensemble of hypotheses can expand the space of functions in such a way that a more accurate approximation of the true function can be obtained. The computational problem arises when it cannot be guaranteed that the learning algorithm will find the best hypothesis within the hypothesis space. That means the task of finding the best hypothesis is computationally intractable and therefore heuristic methods must be employed. These heuristics (such as gradient descent) can get stuck in local minima and hence fail to find the best hypothesis. An ensemble constructed by running a local search from different starting points may provide a better approximation to the true unknown function than any individual algorithm.

We hypothesize that the boosting theory can be transferred into the domain of DIR. In the context of DIR, the hypothesis \( h_n \) is represented by the displacement field \( u_n(x) \) generated by registration \( n \). In this paper we focus on the computational problem, and present an algorithm for boosting deformable image registration (DIRBoost).

2. Methods

DIRBoost is an iterative and adaptive algorithm, similar to AdaBoost (Freund and Schapire, 1997) for classifier boosting. In classifier boosting there is ground truth data available for training and validation of a classifier, and the trained classifier can be applied to automatically classify unseen instances from the same population the classifier was trained on. This is in contrast to DIR where each registration \( n \) requires the establishment of an individual mapping \( u_n \) from physical points of one object to corresponding locations on the other object. We consider the displacement field \( u_n \) of registration \( n \) as the hypothesis \( h_n \). In order to boost a DIR algorithm similarly to classifier boosting, we require an estimate of local registration accuracy. Accordingly, each hypothesis \( u_n \) is weighted by a weight field \( \omega_n \) based on the local error estimate.

2.1. DIR algorithm

Any DIR method that can produce a diffeomorphic mapping can be employed in DIRBoost. We define the interface of a DIR method:

\[
\text{DIR}(u_{n-1}, I_x, I_x),
\]

with \((I_x, I_y)\) the image pair to be registered, \( u_{n-1} \) a displacement field to initialize the deformable registration, and \( u_n \) the obtained displacement field resulting from the current registration. In DIRBoost we distinguish between the base registration (iteration \( n = 0 \)) and the actual boosting registration starting at \( n = 1 \). Hence, the output displacement field of the base registration, \( u_{n-1} \), is the initial input displacement field for boosting.

Next to a DIR method, DIRBoost requires a method for registration error detection, RED.

2.2. Registration error detection

Any RED method that can provide a voxel-wise quality map of image registration can be employed, such as the methods proposed by Muenzing et al. (2012a); Risholm et al. (2013); Datteri and Dawant (2012); Watanabe and Scott (2012). We define the interface of a RED method:

\[
\text{RED}(I_x, I_y, u_{n-1}, M_{\text{RED}}).
\]

with \( RE \) the obtained quality map of image registration, \((I_x, I_y)\) the registered image pair, \( u_{n-1} \) the corresponding mapping, and \( M_{\text{RED}} \) a binary mask on the fixed image that defines the region of interest (ROI), that is the part of the image that is considered for registration and for assessment of registration quality.

2.3. Boosting process

The boosting process starts with the detection of errors of the base registration, based on which the boosting masks are generated to define the region of interest to be boosted. The boosting masks are then employed to focus the registration on erroneous regions by driving the DIR algorithm by the intensity information of the masked regions. The obtained displacements at the masked regions are then merged with the displacements of previous registration. We refer to this step as the hypothesis update. A graphical overview of the boosting process is depicted in Fig. 1. In the following the details of the boosting process are described.

2.4. Weighting – boosting masks

In classifier boosting each ensemble member is weighted by a coefficient \( z_n \) that is usually related to the accuracy of the single hypothesis \( h_n \). 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. Similarly, in DIRBoost we propose to weight each hypothesis/registration depending on its estimated error. However, application of a continuous weighting function –similar to AdaBoost– would require a DIR method that allows weighting of its similarity measure. Most DIR methods however do not implement such weighting of similarity metrics, therefore we opt for a simple binary weighting scheme. We define a weight field \( \omega_n(x) \) such that each voxel \( x \) in the region of interest is assigned the binary weight one or zero, i.e. was previously misaligned or correctly registered. In practice this yields simple binary boosting masks \((M_{\text{Boost}}, M_{\text{MBboost}})\) which can be applied to the input images \((I_x, I_y)\) in order to focus the registration on previously incorrectly mapped voxels: \( u_{n+1} = \text{DIR}(u_{n-1}, I_x, I_y, M_{\text{Boost}}; I_x, I_y, M_{\text{MBboost}}) \). Depending on the particular DIR method and the similarity measure employed, the boosting masks are applied either by simply setting the image intensity of the background (i.e. all voxels outside the masks) to a homogeneous value, or by additionally calculating the similarity metric only on the region defined by the boosting masks.

To that end, we first define an error threshold \( e \) to determine erroneous regions such that all voxels of the quality map where \( RE \geq e \) are considered erroneous mappings whereas the remainder is considered as regions with correct or good alignment. Second, we propose a weighting function, \( \delta_{\text{RED}} \), that spatially expands, \( \otimes \), erroneous regions onto the region with good alignment. We determine the extent of the spatial expansion based on the spatial accuracy of the RED method. The weighting function is applied to the moving image domain via the inverse transformation \( T_{n+1}^{-1} \), i.e. the binary boosting mask, \( M_{\text{Boost}} \), obtained in the fixed image domain is mapped onto the moving image domain.

\[
M_{\text{Boost}} = (RE \geq e) \otimes \delta_{\text{RED}},
\]

\[
M_{\text{MBboost}} = M_{\text{Boost}} \circ T_{n+1}^{-1}.
\]

Inverse mappings at locations that are deemed erroneous would yield erroneous correspondences in the moving image. Therefore, the bounding of erroneous regions in the fixed image domain by \( \delta_{\text{RED}} \) is important to obtain a valid estimate of the corresponding region in the moving image domain. Furthermore we assume that the DIR algorithm – initialized with a coarse version of the displacement field obtained in the preceding registration – aligns those boundaries similarly in subsequent boosting iterations resulting in smooth transitions between the preceding and the updated deformation field. Moreover, we hypothesize that these boundary
correspondences facilitate the boosting process by (spatially) bounding the registration problem and therefore aiding the DIR algorithm.

2.5. Hypothesis update

DIRBoost is an iterative algorithm, in the sense that each boosted deformation field \( u_{n+1} \) represents an update on the previous deformation field \( u_n \). Thanks to the binary weighting scheme the update is simply an addition of \( u_{n+1} \) and \( u_n \) which are masked by \( a_n \) and its complement respectively. However, irregularities in the underlying topology of the updated field should be avoided. For that reason a regularization step \( \Phi_{\text{Diff}} \) is included to ensure that the updated field is diffeomorphic. We define the hypothesis update:

\[
\begin{aligned}
u_{n+1} &\leftarrow \Phi_{\text{Diff}}(u_{n+1} \cdot a_n + u_n \cdot a_n^c).
\end{aligned}
\]

Upon assuming that the DIR algorithm generates folding-free deformation fields, only the join borders are prone to irregularities, which can be corrected locally using the following methods. Modat et al. (2010b) propose a regularization scheme based on a B-Spline transformation model which can be applied after registration. It can be considered a correction phase where B-Spline nodes related to foldings are re-positioned such that foldings in the deformation field are avoided. Muenzing et al. (2012c) propose a post-registration regularization method that is based on a fast B-Spline reconstruction scheme. Folding regions are corrected by substituting the field locally by an interpolation from the uncorrupted vicinity of the field; by selecting a particular order of B-Spline the smoothness of the field can be controlled. Alternatively, the regularization \( \Phi_{\text{Diff}} \) itself can be considered as a registration problem. Yushkevich et al. (2012) describe such a regularization scheme for correspondence fusion in groupwise registration. The regularization is formulated as a registration problem where a non-diffeomorphic mapping is to be approximated by a diffeomorphic mapping. The source image is defined by the identity transformation, and the joined deformation field forms the target image. A transformation in diffeomorphic space is then obtained by standard registration tools (Avants et al., 2008) with the mean square difference metric as similarity measure.

2.6. Stopping criterion

We propose a heuristic convergence criterion that allows ending DIRBoost before reaching \( N \) iterations. It is based on the relative amount of registration errors,

\[
EpV^* = \frac{\sum \left( \mathbf{R}^* \geq \epsilon \right)}{\sum M_{\text{ROI}}},
\]

that is the percentage of the initial ROI volume which is deemed erroneous. Iterating is stopped when there is no improvement in two consecutive iterations.
\[ \text{Stop} \rightarrow (EpV^n_s \geq EpV^{n-1}_s) \land (EpV^{n-1}_s \geq EpV^{n-2}_s). \]

When the stopping criterion is met at iteration \( n = s \), iteration \( n = s - 2 \) is chosen as final iteration, and with \( h_n = u_{s-1} \) (cf. Algorithm 1), the final hypothesis \( h_0 = u_{s-1} \). Alternatively, a fixed number of \( N \) iterations can be computed and then the final hypothesis \( h_N \) is selected according to iteration \( n = f \) with the smallest error \( EpV^n_f \). In general, the final hypothesis \( h_0 \) is generated by the hypotheses sequence \( H_{n=1} = \Phi_{\text{Diff}}(u_{s-1} ; a_n + u_n ; a_n^c) \).

**Algorithm 1. DIRBoost**

**Input:**
- \( l_c, I_M \), and initial deformation field \( u_1 \)
- binary masks \( M_{\text{ROI}}, M_{\text{ROI}} \) defining the region of interest
- registration error detection method RED
deforable image registration algorithm DIR
integer \( N \) specifying maximum number of iterations

**Initialize** the boosting masks:
\[
M_{\text{Boost}} = M_{\text{ROI}}
\]
\[
M_{\text{AllBoost}} = M_{\text{ROI}}
\]

**Do for** \( n = 1, 2, \ldots, N \)
1. Set weight field (Section 2.4)
\[
a_n(x) = \begin{cases} 
1 : & M_{\text{ROI}}(x) = 1 \\
0 : & \text{otherwise}
\end{cases}
\]
2. Call DIR algorithm (Section 2.1)
\[
u_{n-1} = \text{DIR}(u_{n-1} ; l_c, l_M, M_{\text{ROI}}, M_{\text{ROI}});
\]
get back hypothesis \( h_n : u_{n-1} \).
3. Update hypothesis (Section 2.5)
\[
u_n = \Phi_{\text{Diff}}(u_{n-1} ; a_n + u_n ; a_n^c),
\]
where \( \Phi \) denotes the Hadamard product with entry-wise multiplications of field vectors \( u \) and \( a \) of the binary complement of \( a \), and \( \Phi_{\text{Diff}} \) represents a regularization such that the updated field is diffeomorphic.
4. Call RED to estimate the registration error (Section 2.2)
\[ RE^n = \text{RED}(l_c, l_M, u_{n-1}, M_{\text{ROI}}), \]
where \( RE^n \) denotes a quality map of the registration.
5. Generate boosting masks (Section 2.4)
\[
M_{\text{Boost}} = (RE^n > \varnothing) \odot \delta
\]
\[
M_{\text{AllBoost}} = M_{\text{Boost}} \circ T_{n-1}^{-1}
\]
where \( \odot \delta \) denotes a spatial expansion of the region \( (RE^n > \varnothing) \) deemed erroneous.
6. Stop iterating (Section 2.6)
if convergence criterion is met.

**Output** hypothesis:
The final hypothesis \( h_f \) is determined either by a convergence criterion or by selection of the best hypothesis of a fixed number of \( N \) iterations, both based on quality estimates \( RE^n \) of corresponding registration.

### 3. Experiments

DIRBoost is a wrapper algorithm, it employs a particular registration method to generate boosted registrations. Further, DIR-Boost requires a method for registration error detection (RED). Both the DIR and RED method are interchangeable components of DIRBoost. In the following we utilize a supervised RED method and we demonstrate the application of DIRBoost employing three different registration methods. We validate DIRBoost on three reference datasets of pulmonary images.

#### 3.1. Reference datasets

##### 3.1.1. NELSON

Five patients (male, ages 51–75 yrs) were chosen randomly from the NELSON lung cancer screening database (Xu et al., 2006), each with a computed tomography scan of breath-hold full inspiration and breath-hold forced expiration, made in the same session. The scans were acquired with a 16 detector MSCT scanner (Philips Brillance 16P, Philips Medical Systems, Cleveland, OH, USA), without contrast. Depending on the body weight (<50, 50–80 and >80 kg) the kVp settings were 80–90, 120 and 140 kVp, respectively. The inspiration scan was created using a low-dose protocol (30 mAs) whereas the expiration scan was ultra-low-dose (20 mAs), both with a slice spacing of 0.70 mm and pixel spacing between 0.63 mm and 0.77 mm, without contrast. Reference landmark matchings were manually established by the first author using specifically developed validated software (Murphy et al., 2011a). For each scan pair a set of 100 reference landmarks \( L_f \) was automatically defined on the fixed image (expiration scan), and then matched with the corresponding point \( L_M \) in the moving image (inspiration scan). Lung masks were computed for each scan employing an automatic lung segmentation method (Hu et al., 2001). The scan pairs exhibit on average an increase of the lung volume of about 80% [56,140] from the expiration status to the inspiration status.

##### 3.1.2. COPDgene

DIR-Lab (www.dir-lab.com) at the University of Texas M.D. Anderson Cancer Center in Houston, TX, released recently a reference dataset based on the COPDgene study archive along with large sets of manually identified anatomic landmark pairs (Castillo et al., 2013). We utilize this dataset which comprises 10 breath-hold computed tomography (BH-CT) image pairs, randomly selected from the COPDgene study cases. Each patient received CT imaging of the entire thorax in the supine position at normal expiration and maximum effort full inspiration. CT imaging was performed with a GE VCT 64-slice scanner (GE Healthcare Technologies, Waukesha, WI) with a pitch of 1.375 mm, speed of 13.75 mm per rotation, 120 kVp, 0.5 s per rotation, 400 mA per rotation for inhale BH, and 100 mA per rotation for exhale BH. The images used in this study were reconstructed with a high-resolution (BONE) kernel, with the lung diameter setting the field of view, and with contiguous 2.5 mm slice spacing and in-plane voxel dimensions ranging from \((0.590 \times 0.590)-(0.652 \times 0.652) \text{ mm}^2\). An imaging expert manually identified large sets of anatomical landmark pairs between images, using dedicated in-house software. Estimates of inter- and intra-observer spatial variation in landmark localization were determined by repeated measurements of multiple observers over subsets of randomly selected landmarks. A total of 7298 anatomical landmarks were manually paired between the 10 sets of images. The quantity of landmark pairs per case ranged from 447 to 1172. Repeat registration of uniformly sampled subsets of 150 landmarks for each case yielded estimates of observer localization error, which ranged from a mean
of 0.58 (SD: 0.87) to 1.06 (SD: 2.38) mm. In addition, we computed lung masks for each image employing an automatic lung segmentation method (Hu et al., 2001). The scan pairs exhibit on average an increase of the lung volume of about 62% [26,109] from the expiration status to the inspiration status.

3.1.3. EMPIRE10

EMPIRE10 is a public platform (empire10.isi.uu.nl) for evaluation of methods for pulmonary image registration, designed for the image registration challenge at MICCAI 2010 (Murphy et al., 2011b). The EMPIRE10 dataset consists of 30 intra-subject thoracic CT image pairs gathered from several sources, acquired by different types of CT scanners and imaging protocols. The image data is categorized in the following six types:

- breathhold inspiration scans
  - 8 image pairs: # 2,3,9,11,15,19,22,27
- breathhold inspiration and expiration scans
  - 8 image pairs: # 1,7,8,14,18,20,21,28
- 4D scans of regular breathing
  - 4 image pairs: # 13,16,17,23
- ovine lung scans
  - 4 image pairs: # 4,10,24,29
- contrast - non-contrast scans
  - 2 image pairs # 6,26
- artificially warped scans
  - 4 image pairs # 5,12,25,30

In addition to the cropped scan data, lung masks are provided for every scan. See (Murphy et al., 2011b) for a full description of the dataset. For evaluation of computed image registrations, the generated displacement fields have to be submitted to the provider of the platform, where the registration results are evaluated with respect to alignment of the lung boundaries, alignment of the major fissures, correspondence of matched point pairs, and analysis of singularities in the deformation field.

3.2. Registration algorithms

We demonstrate the DIRBoost algorithm on three different registration methods, viz. ANTS, NiftyReg, and 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 (Klein et al., 2009; Murphy et al., 2011a). 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 (Murphy et al., 2011b). We employ these registration configurations to obtain optimized base registrations.

3.2.1. ANTS

ANTS (Avants et al., 2013) is an open source software package (www.picsl.upenn.edu/ANTS) built on the ITK (Insight Segmentation and Registration Toolkit) framework (National Library of Medicine, 2013). ANTS comprises a suite of tools for image normalization and template building, in particular a diffeomorphic registration method with symmetric normalization (Avants et al., 2008).

Base registration. The computation of the ANTS base registration is based on the registration set-up and parametrization proposed in Song et al. (2010). 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 (Avants et al., 2008). 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 the diffeomorphic registration, gradient descent is used in the optimization. Convergence of 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 to avoid trapping in a local minimum.

We adapted the proposed multi-resolution setting (Song et al., 2010) to further reduce computation time, conducting diffeomorphic registrations only until half the original image resolution, which setting we denote in the sequel by the acronym ANTS2. We compared the ANTS2 setting with the full-resolution approach denoted by ANTS and found no substantial differences in the results.

DIR set-up for boosting. We use the ANTS2 parameterization 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 deformation field; we therefore implement the indirect initialization approach described in Section 3.2.4. For regularization of the deformation field in the hypothesis update step, an external post-registration regularization method (cf. Section 2.5) is employed.

3.2.2. NiftyReg

The NiftyReg (sourceforge.net/projects/NiftyReg) registration package (version 1.3) contains a global and a 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 Modat et al. (2010a) but perform all registrations until convergence. The proposed configuration makes use of lung masks to focus the registration on the lungs. The base registration is a multi-resolution approach consisting 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.

DIR set-up for boosting. We use the same parameterization 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.

3.2.3. 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 (Glocker et al., 2008). 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 Glocker et al. (2010). 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 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 used the same parameterization employed in the base registration but require 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 deformation field. We therefore implement the indirect initialization approach described in Section 3.2.4. For regularization of the deformation field in the hypothesis update step, we employ an external post-registration regularization method based on fast B-Spline reconstruction (cf. Section 2.5).

3.2.4. 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 \( u_0 \) is smoothed with a Gaussian kernel with \( \sigma \) equal to the coarsest voxel spacing applied in the registration pipeline. Second, the moving image \( I_{R0} \) is warped using the smoothed displacement field \( u_0 \), yielding the registered image \( I_{R0} \). Third, image registration is performed employing the image pair \((I_R, I_{R0})\) instead of \((I_R, I_{R0})\). Fourth, the obtained displacement field \( u_{R0} \) is composed with \( u_0 \) to account for the initial warp of \( I_{R0} \). In order to boost registrations optimally, the DIR algorithm should have access to the original image information of \( I_{R0} \), which might be distorted in \( I_{R0} \) beyond recognition and therefore might prevent the DIR algorithm from recovering the underlying deformation between \( I_R \) and \( I_{R0} \). The use of \( I_{R0} \) instead of \( I_R \) as surrogate initialization is crucial to alleviate possible image distortions.

3.3. Registration error detection

We employ an automatic method for quality assessment of medical image registration (Muenzing et al., 2012a) 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.

3.3.1. 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 \( L_0 \) on the fixed image \( I_0 \) and their corresponding location \( L_{M0} \) in the moving image \( h_0 \). The set of reference image registrations consists of transformations \( T \) obtained by the particular registration method on the particular image data, on which registration errors shall be detected. The set of statistical image features \( F(S) \) contains for each landmark of \( L_0 \) the corresponding feature values that are computed based on different image feature types. Gaussian, correlation, and entropy features are calculated from the intensity images \( (I_0 \) and \( I_R) \), and deformation features are computed on the transformations \( T \).

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 \( T(L_0) \) and the corresponding reference matching \( L_{M0} \). To automatically classify a previously unseen registration, first, a large set of landmarks \( L_0 \) is automatically detected on corresponding \( I_0 \). Second, image features are extracted according to \( F(S) \) for each landmark of \( L_0 \). 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 \( L_{M0} \), 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 Muenzing et al. (2012a). 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 100 mm. The set of image features described in Muenzing et al. (2012a) has been extended by the following geometrical, deformation-based and cubic subvolume intensity features. Normalized landmark position \( NLP(x) \): physical position normalized by bounding box of ROI in \( M_R \). Distance of landmark to ROI boundary \( DTB(x) \): computed based on distance transformation of \( M_R \) and \( M_M \). Normalized Jacobian map \( NJM(x; \sigma) : \{f(u(x) + G(x; \sigma)) / \sqrt{V} \} \) with \( \sigma \) and \( V \) the volume ratio of the ROI between \( I_R \) and \( I_{M0} \). Mean of absolute and squared intensity difference \( MID(x; \sigma) \), \( MSD(x; \sigma) \). In addition, texture features \( GLCM(x; \sigma) \) based on gray-level co-occurrence matrices (GLCM) are included for application to 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.

3.3.2. Training and application of RED

In the following a distinction is made between the NELSON and COPDgene reference datasets, where we have reference landmark
matchings available for training and validation, and the EMPIRE10 reference dataset, where a dedicated evaluation of registrations is provided through the online platform.

For the NELSON and COPDgene dataset, we computed reference image registrations for each application of RED, that is, for each reference image data and for each registration method. Then, for each intended application of RED, the specific learning dataset \( S \) was established and leave-one-scanpair-out cross-validation was performed. This means, each learning dataset \( S \) is split into \( k \) mutually exclusive subsets so that \( S \) contains landmark-related data of scan pair \( k \) exclusively. See Muenzing et al. (2012a) for details on the optimization and selection process of multi-feature-classifier models. We set the quality categories \( \text{CA}, \text{PA}, \text{WA} \), based on the landmark registration error \( \text{LRE} \), to the following values: \( \text{CA} = \{ \text{LRE} \leq 2 \text{ mm} \} \), \( \text{PA} = \{ 2 < \text{LRE} < 5 \text{ mm} \} \), \( \text{WA} = \{ \text{LRE} \geq 5 \text{ mm} \} \). However, different values could be chosen to define registration quality and registration errors. The random forests (RF) classifier was used, employing 1000 trees and the default feature subsampling setting. We limited the size of the feature subsets so that the AUC 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. The classification accuracy predicted by cross-validation is on average 80% on the NELSON dataset (DROP: 76% [0.88,0.85], NiftyReg: 81% [0.77,0.89], ANTS: 82% [0.78,0.84]), and on average 77% on the COPDgene dataset (DROP: 72% [0.91,0.85], NiftyReg: 75% [0.90,0.79], ANTS: 85% [0.89,0.79]). Values in braces denote corresponding AUC of domain 1 and domain 2 respectively.

For application of the RED classifier we maintain the cross-validation set-up, that is, registration quality estimates on scan pair \( k \) are based on a training dataset \( S \setminus S_k \), including landmark related data from all scan pairs but \( k \). First, a set of statistical landmarks \( l_k \) is automatically detected on the fixed image \( l_k \) of scan pair \( k \), using the landmark detection method described in Murphy et al. (2011a). To obtain a spatially dense estimation of registration quality, we choose parameter settings (minimum distance to neighboring landmarks, sampling density) of the landmark detection method such that theoretically 1,000 landmarks per dataset can be determined (upper bound). In practice we expect fewer landmarks to be detected, however. Only about 500–600 landmarks could be determined, which can be explained by the relatively strong image noise, and the occurrence of emphysema appearing as diffuse lesions or air-filled bullae (end-stage emphysema). Second, statistical image feature \( f_S \) are extracted at landmarks \( l_k \), and classified by the trained RED classifier. Finally, the RED system returns an estimate of registration quality in the form of a quality map \( \text{RE}^* \), where \( \text{RE}^* \geq \epsilon \) marks registration errors with \( \epsilon = 5 \text{ mm} \) based on the quality category WA defined in the training procedure.

The EMPIRE10 challenge does not provide training data for independent evaluation, needed for automatic training or manual adjustment of registration methods to particular application data. In order to properly train a RED classifier for the mixed image data of the EMPIRE10 set, we compose a reference set consisting of the NELSON and COPDgene breathhold references, and an additional reference of 5 follow-up scan pairs of the NELSON study archive, which we had established earlier (Muenzing et al., 2012a). Similarly to NELSON and COPDgene, reference image registrations were obtained, a learning dataset \( S \) was established, and the classifier was cross-validated. For application however, the RED classifier was trained on the entire learning dataset \( S \) comprising breath-hold and follow-up scan pairs.

### 3.4. Weighting function

The purpose of the weighting function \( \delta_{\text{RED}} \) is the spatial expansion of erroneous regions onto well aligned regions (cf. Section 2.4). The extent of the spatial expansion needed is dependent on the spatial accuracy of the RED method. For the employed landmark-based RED method, the spatial accuracy is determined by the distance between neighboring landmarks \( l_k \). Although the landmarks \( l_k \) are aimed to be equally distributed throughout the lung volume, the landmark density varies locally. The averaged mean landmark distance and the overall minimum and maximum landmark distance is on the NELSON dataset 15.47 [12.03, 26.39] mm, and on the COPDgene dataset 13.51 [10.00, 29.73] mm. To optimally account for the possible variations, we opt to implement \( \delta_{\text{RED}} \) based on the employed RED method. The Voronoi decomposition performed in RED results in Voronoi cells centered at the position of the landmarks \( l_k \). We implement \( \delta_{\text{RED}} \) therefore as the spatial expansion of WA regions halfway onto their neighboring Voronoi cells (cf. Section 3.3.1).

### 4. Evaluation and results

We validated the DIRBoost algorithm employing three different DIR methods on three different reference datasets of intra-patient lung CT registration. We calculated the following measures for the evaluation of DIRBoost on the reference datasets. The mean and maximum landmark registration error (LRE) was calculated, and to assess the amount of larger misalignments, the percentage of landmarks with LRE larger than 5 mm, 10 mm, and 20 mm. In addition to the landmark-based measures, which primarily evaluate the registration accuracy of the interior of the lungs, we assess

<table>
<thead>
<tr>
<th>DIR</th>
<th>LRE (mm)</th>
<th># landmarks (%)</th>
<th>Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Max</td>
</tr>
<tr>
<td>(a) NELSON</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANTS</td>
<td>1.68</td>
<td>3.0</td>
<td>16.5</td>
</tr>
<tr>
<td>Base</td>
<td>2.16</td>
<td>5.3</td>
<td>16.5</td>
</tr>
<tr>
<td>Boosted</td>
<td>2.61</td>
<td>5.3</td>
<td>16.5</td>
</tr>
</tbody>
</table>

| (b) COPDgene|          |     |     |    |     |     |       |            |
| ANTS        | 2.42     | 3.5 | 18.7| 13.0| 7.0 | 1.2 | 99.2   |            |
| Base        | 2.71     | 3.5 | 18.9| 13.5| 7.0 | 1.2 | 98.8   |            |
| Boosted     | 2.79     | 3.5 | 18.9| 13.5| 7.0 | 1.2 | 98.8   |            |

| NiftyReg    |          |     |     |    |     |     |       |            |
| Base        | 3.14     | 3.2 | 16.9| 18.7| 4.8 | 0.5 | 98.5   |            |
| Boosted     | 2.19     | 5.6 | 16.9| 18.7| 4.8 | 0.5 | 98.5   |            |

| DROP        |          |     |     |    |     |     |       |            |
| Base        | 6.92     | 6.3 | 31.6| 42.6| 22.9| 7.2 | 98.2   |            |
| Boosted     | 4.68     | 4.1 | 21.5| 29.5| 11.6| 1.5 | 98.2   |            |
the boundary alignment by the Dice overlap between the fixed and registered lung masks. Further, to quantify possible irregularities of the displacement field, the number of foldings was computed, that is the number of voxels within the fixed lung mask where the determinant of the Jacobian matrix is non-positive. Cross-validation (leave-one-scanpair-out) was used to avoid classification bias of the RED method. Moreover, the employed DIR algorithms do not involve any landmark matching, the optimization is purely intensity-based.

Table 1 summarizes the evaluation results on the NELSON and COPDgene reference datasets. It lists the averaged results of the base and boosted registrations of ANTS2, NiftyReg, and DROP, as well as the ANTS gSyN benchmark registrations denoted by ANTS. None of the registrations contained any singularity or folding irregularities.

Fig. 2 depicts the progress of the boosting based on the error estimate $E_{PV}$ of the RED method. Note that the boosted registrations listed in Table 1 are all selected based on the convergence criterion (Section 2.6). DIRBoost registrations converged on the NELSON dataset on average within four, three, and seven iterations for ANTS2, NiftyReg, and DROP; and within five, three, and seven iterations on the COPDgene dataset.

Table 2 summarizes the EMPIRE10 evaluation results provided via the EMPIRE10 online platform. The EMPIRE10 evaluations consist of boundary, fissure, landmark, and singularity scores. The lung boundary and fissure scores represent alignment penalties based on semi-automatic reference segmentations of the lung boundary and the major fissures. The singularity score is the relative number of voxels with foldings within the lungs. The landmark score is based on the averaged Euclidean distance between registered reference landmarks and the corresponding manual matchings. Note that the EMPIRE10 evaluation measures involve some tolerance to account for possible inaccuracies of the ground truth. See (Murphy et al., 2011b) for a detailed description of the evaluation methods. Table 2(a) summarizes the EMPIRE10 evaluation results based on the entire set of 30 registrations. 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. See Table 2 for the image pair ID and Section 3.1.3 for the corresponding image data category. Table 2(b) lists the EMPIRE10 evaluation results of the boosted registrations exclusively.

Overall, DIRBoost could improve registrations of each of the DIR methods on all datasets. On the NELSON and COPDgene dataset the absolute reduction of registration errors (misalignments larger than 5 mm) varied in the range $0.001$ to $0.013$, assessed based on reference landmark correspondences. Furthermore we tested the statistical significance of the error reduction by comparing the LRE before and after boosting of all landmark matchings with initial misalignment larger 5 mm, separately for each DIR method and reference dataset. Paired t-tests revealed a statistically significant reduction of registration errors ($p < 0.02$). The reduction of registration errors yielded improvements of the average LRE in the range $0.1$ to $0.24$ mm, which accounts for about 5–32% improvement of the registration accuracy. In general, the COPDgene scan pairs appear to be more difficult to register and therefore show more potential for boosting than the NELSON scan pairs.
The EMPIRE10 evaluation results show that DIRBoost generated transformations without any singularities or foldings. Overall, boosted registrations are superior to base registrations, according to the EMPIRE10 ranking scheme. The registration improvements due to boosting are more distinct in Table 2(b) which solely summarizes the evaluation results of those registrations that are scaled to display relevant digit positions.

5. Discussion

5.1. Comments on registration results

The difficulties underlying the image data are the large volume differences between scan pairs along with strong noise (non-uniformly distributed) which can result locally in large distortions during the registration process. Boosting of ANTS2 yielded only slight improvements on the NELSON data, however clearly improved the COPDgene registrations, yielding comparable registration accuracy on both reference standards ultimately. Further, DIRBoost improved the NiftyReg registrations to a level comparable to the ANTS benchmark registrations, consistently on both reference standard datasets. The effectiveness of DIRBoost is also demonstrated with DROP. Although DIRBoost significantly improved the DROP registrations, the obtained registration accuracy, however, is not on a par with NiftyReg or ANTS. The results of boosted DROP registrations illustrate the characteristics of DIRBoost: it aims to improve non-systematic errors, and does this within a few iterations, however the overall registration capability clearly appears to be determined by the employed DIR method.

The EMPIRE10 evaluations Table 2 show that DIRBoost could improve on each registration method. 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 (Song et al., 2010; Modat et al., 2010a). 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.

5.2. Impact of RED method on DIRBoost

DIRBoost is based on the assumption that it is possible to automatically assess local registration quality. We employ our previously developed supervised pattern recognition method for assessment of registration quality, however, DIRBoost is methodologically not restricted to this particular registration quality assessment. Further, the manual matching required for the training of RED is only necessary one time for a particular application.

The employed supervised RED system achieves an overall classification accuracy of about 78%. Although this constitutes a good classification result, it is certain that misclassifications occur, especially at class borders, i.e., between quality categories CA and PA, and PA and WA. To estimate the possible impact of errors of the RED method, we exemplarily analyzed the boosted ANTS2 registrations of the COPDgene reference standard dataset. We compared the boosting mask of the final hypothesis with the LRE of the reference landmark matching. The boosting masks captured on average 89%, 95%, and 100% of the landmarks with remaining LRE >5 mm, >10 mm, and >20 mm, respectively. Although this analysis on the final hypothesis is merely an indication for the overall impact during the boosting progress, it suggests that more accurate RED estimates could yield better DIRBoost results.

In Fig. 2(d) of the ANTS2 – COPDgene registrations there is one particular registration that stands out (depicted by circles). Its mean LRE is 6.9 mm with 38% of the reference landmarks misaligned >5 mm. After boosting, the mean LRE is reduced to 2.0 mm with 7% of landmarks misaligned >5 mm. This improvement is however not properly reflected by error estimates EpV of the RED method. We inspected visually the boosted registration and found that the registered image appears blurry and irregular in comparison to the other ANTS2 registrations, probably due to the strong noise present in the scan pairs. Although the RED method overestimated the LRE, its relative error estimates seemed to be sufficient for boosting.

The evaluation of DIRBoost on the EMPIRE10 dataset shows that the employed RED method detected registration errors on all relevant registrations but scan pair 10 (cf. detailed result list online). Scan pair 10 is one of four ovine registrations included in the EMPIRE10 dataset. We expected suboptimal error estimates on these cases because the RED method was trained on registrations of the RED method. We inspected visually the boosted registration and found that the registered image appears blurry and irregular in comparison to the other ANTS2 registrations, probably due to the strong noise present in the scan pairs. Although the RED method overestimated the LRE, its relative error estimates seemed to be sufficient for boosting.

The evaluation of DIRBoost on the EMPIRE10 dataset shows that the employed RED method detected registration errors on all relevant registrations but scan pair 10 (cf. detailed result list online). Scan pair 10 is one of four ovine registrations included in the EMPIRE10 dataset. We expected suboptimal error estimates on these cases because the RED method was trained on registrations of human scan pairs only. On the ovine cases only about 100 landmarks were detected, compared to about 300–600 landmarks on the human cases.

5.3. Why/how boosting of DIR works

There are two key properties of proposed DIRBoost algorithm: 1. Focus of the similarity measure on incorrectly registered regions, thereby refining the similarity measurement of the DIR method and avoiding possibly distracting image information contained in
surrounding correctly registered regions. 2. Definition of a spatial subspace of possible transformations by employing masks (bounds) on the fixed and the moving image domain. Both properties indirectly define a search subspace on the optimization problem (underlying the registration), and we hypothesize that this bounding enables the iterative optimization process to overcome local optima and/or converge to better local solutions. In Fig. 2, the boosting progress appears to be oscillating at the end of the boosting process. We assume that this is related to the issue of locality, i.e. finding the optimal region size to be boosted: too much focus on structural sparse or noisy image regions yields little image information and therefore unstable similarity measurements. More generally, if true correspondence is missing (e.g. tumor growth or resection) or image acquisition/information is insufficient (e.g. noise, emphysema) then larger regions (containing more globally robust correspondences) are needed to overcome locally insufficient similarity measurements. In analogy to classifier boosting, we assume that DIR boosting is most efficient for moderately complex but fast algorithms, such as the NiftyReg and DROP algorithm. Both algorithms employ a global similarity measure along with a fast optimization scheme. On the other hand, less performance improvement is expected from boosting of more complex registration methods, such as the ANTS gSyn method. The use of boosting for complex and computationally expensive registration methods is therefore less attractive.

5.4. Computational considerations

Boosting registrations are performed on subregions and require therefore in general less time until convergence than the base registrations. In our experiments, the implementation of DIRBoost, and in particular the RED method employed, have not yet been optimized for computational speed. However, we assume that the time requirements of an optimized implementation will be negligible compared to the computation times of most DIR algorithms. Therefore, the upper bound of the overall computation time for boosting might be roughly estimated as the computation time of the base registration x number boosting iterations. The average number of boosting iterations on the NELSON dataset is four, three, and seven iterations for ANTS2, NiftyReg, and DROP; and five, and three, and seven iterations on the COPDGene dataset (Section 4). Considering the achieved accuracy improvements, the average number of boosting iterations needed, and the computation time requirements of the registration software; DIRBoost with the GPU implementation of the NiftyReg algorithm appears to yield the best computational performance. It should be noted, that the boosting performance of the ANTS2 and the DROP registration method might be underestimated. Both software implementations (ANTS and DROP) do not provide the option to initialize a registration with a deformation field, and therefore an indirect initialization approach (Section 3.2.4) has been used as substitute. The additional image resampling step and possibly remaining image distortions might negatively affect the boosting registrations, and might consequently yield limited overall performance of the boosting approach for those algorithms.

5.5. Future work

DIRBoost is based on the computational problem stated in Section 1.2, i.e. DIRBoost could possibly boost any DIR algorithm with the given properties of the employed DIR algorithm such as similarity metric and transformation model. Those possibly limiting properties might also be addressed within DIRBoost, similar to modified versions of AdaBoost, such as feature-selective AdaBoost (Tieu and Viola, 2004). That means, transferred to DIRBoost, that the DIR algorithm could be configured with different similarity metrics, so that in every iteration the DIR algorithm is applied separately for each similarity metric and then the resulting registration with the best accuracy is chosen to update the current hypothesis (similarity measure-selective DIRBoost).

In classifier boosting different weighting functions were investigated (see e.g. Meir and Rätsch (2003)), therefore another point for further research is the weighting function applied in DIRBoost. The specification of a weighting function can also be related to the issue of locality discussed in Section 5.3.

In this work we have applied DIRBoost exclusively to the registration of intra-subject lung CT scans, however, we believe that the proposed basic concepts are applicable to other applications. We assume the method to work for any registration problem (anatomy, modality) for which local misregistration can be detected.

6. Conclusion

We presented a novel algorithm for boosting of deformable image registration (DIRBoost). Similar to classifier boosting, DIRBoost iteratively employs a registration method and adaptively focuses the registration on remaining errors. In DIRBoost we aim to reduce non-systematic registration errors, thereby obtaining more robust registrations and overall improved registration results. We validated DIRBoost on three different DIR algorithms (ANTS gSyn, NiftyReg, and DROP) on three independent reference datasets of pulmonary intra-subject images. DIRBoost consistently and significantly reduced registration errors yielding an improvement of the registration accuracy by about 5–30% depending on the dataset and the registration algorithm employed. The proposed boosting approach is novel to image registration, and the boosting performance obtained on pulmonary image data shows promise for application of DIRBoost to other registration problems.

References


National Library of Medicine, 2013. ITK, Insight Segmentation and Registration Toolkit.


