

RESEARCH ARTICLE

Comprehensive evaluation of ten deformable image registration algorithms for contour propagation between CT and cone-beam CT images in adaptive head & neck radiotherapy

Xin Li¹, Yuyu Zhang², Yinghua Shi², Shuyu Wu¹, Yang Xiao¹, Xuejun Gu³, Xin Zhen^{1*}, Linghong Zhou^{1*}

1 Department of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, China, **2** Department of Radiotherapy Oncology, The First Hospital of Jilin University, Changchun, Jilin, China, **3** Department of Radiotherapy Oncology, The University of Texas, Southwestern Medical Center, Dallas, Texas, United States of America

* xinzhen@smu.edu.cn (XZ); smart@smu.edu.cn (LZ)



OPEN ACCESS

Citation: Li X, Zhang Y, Shi Y, Wu S, Xiao Y, Gu X, et al. (2017) Comprehensive evaluation of ten deformable image registration algorithms for contour propagation between CT and cone-beam CT images in adaptive head & neck radiotherapy. PLoS ONE 12(4): e0175906. <https://doi.org/10.1371/journal.pone.0175906>

Editor: Qinghui Zhang, North Shore Long Island Jewish Health System, UNITED STATES

Received: October 21, 2016

Accepted: April 2, 2017

Published: April 17, 2017

Copyright: © 2017 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by The National Natural Science Foundation of China (No 81301940 and No 81571771, <http://www.nsf.gov.cn/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Deformable image registration (DIR) is a critical technic in adaptive radiotherapy (ART) for propagating contours between planning computerized tomography (CT) images and treatment CT/cone-beam CT (CBCT) images to account for organ deformation for treatment re-planning. To validate the ability and accuracy of DIR algorithms in organ at risk (OAR) contour mapping, ten intensity-based DIR strategies, which were classified into four categories—optical flow-based, demons-based, level-set-based and spline-based—were tested on planning CT and fractional CBCT images acquired from twenty-one head & neck (H&N) cancer patients who underwent 6~7-week intensity-modulated radiation therapy (IMRT). Three similarity metrics, i.e., the Dice similarity coefficient (DSC), the percentage error (PE) and the Hausdorff distance (HD), were employed to measure the agreement between the propagated contours and the physician-delineated ground truths of four OARs, including the vertebra (VTB), the vertebral foramen (VF), the parotid gland (PG) and the submandibular gland (SMG). It was found that the evaluated DIRs in this work did not necessarily outperform rigid registration. DIR performed better for bony structures than soft-tissue organs, and the DIR performance tended to vary for different ROIs with different degrees of deformation as the treatment proceeded. Generally, the optical flow-based DIR performed best, while the demons-based DIR usually ranked last except for a modified demons-based DISC used for CT-CBCT DIR. These experimental results suggest that the choice of a specific DIR algorithm depends on the image modality, anatomic site, magnitude of deformation and application. Therefore, careful examinations and modifications are required before accepting the auto-propagated contours, especially for automatic re-planning ART systems.

Introduction

Intensity-modulated radiation therapy (IMRT) plays a critical role in the management of head and neck (H&N) cancer patients [1]. IMRT can maximize tumor coverage and/or sparing of organs at risk (OARs) by generating a steep dose gradient and, thus, leads to a potential increase in the therapeutic index [2, 3]. However, it does not take into account fractional anatomical changes, such as tumor shrinkage, nodal/glandular volume change, weight loss and geometric variations, etc., during a typical 6~7-week treatment course [4, 5]. Adaptive radiation therapy (ART) [6] is a possible solution to overcome these limitations by adapting temporal changes in anatomy with daily imaging in the radiotherapy process [7–9]. In a typical online ART process, a cone-beam computed tomography (CBCT) scan is usually performed to obtain three-dimensional patient information prior to treatment, and a deformable image registration (DIR) technique is utilized to establish the correspondence between voxels in the planning computed tomography (CT) and CBCT scans for 1) propagating the contoured region of interest (ROI) from one image to another [10, 11] and 2) deforming the re-planned dose maps to integrate the accumulated dose administered to the patient [12, 13].

Effective and reliable ART relies upon accurate DIR-propagated ROIs, and thus, it is necessary to verify the accuracy of the available DIR algorithms for use in a clinical setting. There are a variety of studies that have assessed the quality of CT-CT DIR algorithms in H&N cancer patients; for instance, Castadot *et al.* [14] compared twelve voxel-based DIR strategies in ROI propagation performance using expert physician-drawn contours as benchmarks. Hardcastle *et al.* [15] investigated the clinical utility of two DIR algorithms for ROI propagation using CT image data from five institutions. They reported good anatomical agreements for OARs but claimed that the propagated target structures need to be thoroughly reviewed before clinically use. Mohamed *et al.* [16] developed a quality assurance workflow for the quantitative assessment of four different DIR techniques used for H&N radiation therapy-simulation CT with diagnostic CT co-registration. A total of 2720 anatomic and 50 tumor/nodal ROIs were used for validation, and improved performance with DIR over rigid registration was observed. Limited studies, however, have focused on the comparison of CT-CBCT DIR algorithms in assessing H&N patient data. Though a number of CT-CBCT DIR methods for different clinical sites have been developed [11, 17–23], it is acknowledged that some DIR algorithms may be more suitable for specific anatomies and image modalities due to the mathematical basis of the algorithms [24]. In this sense, it is beneficial to assess the clinical acceptability of the available DIR algorithms to reduce the time and resources required for contour review and correction in the CBCT-based ART process [15].

The aim of this study was to develop a strict and objective methodology to compare different DIR strategies for contour propagation from the planning CT (PCT) to the treatment CBCT of H&N cancer patients in an adaptive radiotherapy setting.

Materials and methods

Patient cohort

Seven female and fourteen male patients diagnosed as nasopharyngeal cancer (NPC) were included in this study. Details of the evaluated patients are listed in Table 1. All patients were treated by IMRT with fractional CBCT for setup error correction. Immobilization was necessary and was achieved by a customized thermoplastic mask during the acquisition of the PCT and CBCT. The PCT images were reconstructed with a resolution of $1.60 \times 1.60 \times 5 \text{ mm}^3$ along the x, y and z directions, respectively. The CBCT images at a resolution of $0.51 \times 0.51 \times 1.99 \text{ mm}^3$ were acquired before the treatment. The size of axial images for both PCT and CBCT was

Table 1. Patients' characteristics.

Patient #	Gender	Age	Histology	TNM stage	Concurrent CBCT
1	F	44	NPC	T4N1M0 IVA	Yes
2	F	45	NPC	T2N3M0 IVA	Yes
3	F	48	NPC	T3N3M0 IVA	Yes
4	F	59	NPC	T2N2M0 III	Yes
5	M	41	NPC	T1N2M0 IVA	Yes
6	F	67	NPC	T4N3M1 IVB	Yes
7	M	64	NPC	T3N1M0 III	Yes
8	M	50	NPC	T4N1M0 IVA	Yes
9	M	39	NPC	T4N2M0 IVA	Yes
10	M	55	NPC	T2N3M0 IVA	Yes
11	M	39	NPC	T3N2M0 III	Yes
12	M	60	NPC	T2N0M0 II	Yes
13	F	55	NPC	T2N2M0 IVA	Yes
14	M	63	NPC	T2N1M0 IVA	Yes
15	M	46	NPC	T1N2M0 III	Yes
16	F	48	NPC	T2N2M0 III	Yes
17	M	52	NPC	T4N2M0 IVA	Yes
18	M	48	NPC	T4N1M0 IVA	Yes
19	M	37	NPC	T4N1aM0 IVB	Yes
20	M	40	NPC	T4N1M0 IVA	Yes
21	M	67	NPC	T4N1M0 IVA	Yes

<https://doi.org/10.1371/journal.pone.0175906.t001>

512×512. There were 89 slices in each CBCT scan, and the number of slices in the PCT images ranged from 60 to 130. A total of 21 PCTs and 129 CBCTs were utilized in this study.

In this study, all the CT and CBCT images were retrospectively collected and anonymized for clinical research purposes. Our study did not involve new CT or CBCT scans of patients and, therefore, did not require approval of the local ethics committee. All patients provided written informed consent for their images to be used for research purposes and to be published.

Delineation of ROIs

OARs including the parotid gland (PG), the submandibular gland (SMG), the cervical vertebra (VTB) and the vertebral foramen (VF), on both PCT and CBCT were manually delineated by an experienced physician using a commercial treatment planning system (TPS, Eclipse 10.0, Varian) and double checked by this same physician three months later to ease intra-observer variations. The contours delineated on the PCT served as the moving contours being deformed to the reference fractional CBCT domain, while the contours delineated on the CBCT images served as the ground truths for contour propagation accuracy validation in this study. We chose the deformable organ PG and SMG instead of the gross target volume (GTV) for this evaluation study for two reasons. First, many pilot studies have revealed that large volume shrinkage or deformation can be seen in the PG and the SMG during H&N radiation treatment [4, 25–36], and thus, contour propagation of the PG and SMG onto the fractional CBCT anatomy is usually necessary for treatment re-planning or planning optimization. Second, GTV in CBCT images is difficult to identify due to the low soft tissue contrast and severe image artifacts present in CBCT. Using GTV with erroneous contouring in CBCT images as the ground truth would bias the evaluation result.

Table 2. Overview of Ten DIR Algorithms.

Class	Algorithm	Toolkit/Implementation Environment	Acronym
Optical flow	Original Horn-Schunck method	DIRART/Matlab	HS
	Combined Horn-Schunck and Lucas-Kanade method	DIRART/Matlab	HSLK
	Free-Form Deformation method	DIRART/Matlab	FFD
Demons	Original Demons method	DIRART/Matlab	OD
	Modified Demons method	DIRART/Matlab	MD
	Symmetric Force Demons method	DIRART/Matlab	SFD
	Double Force Demons method	DIRART/Matlab	DFD
	Deformation with Intensity Simultaneously Corrected	CUDA/GPU	DISC
Level-set	Original Level-Set Motion method	DIRART/Matlab	LS
Spline	B-Spline method	Elastix/C++	BSpline

<https://doi.org/10.1371/journal.pone.0175906.t002>

Image preprocessing

Before performing the rigid and deformable registration, all the images were pre-processed using an open source software 3D Slicer (version 4.3.1, <http://www.slicer.org>) [37]. Unnecessary image content such as the treatment table and the thermoplastic mask were manually segmented out from all the images. Because the superior-inferior (S-I) patient volume range in the PCT was generally larger than that in the CBCT, in order to unify the image size and save computational time, the PCT was cropped and re-sampled to match the dimension and resolution of the CBCT after rigid registration, which was also done in 3D Slicer. Hence, the image resolution for both the PCT and CBCT images after rigid registration was 256×256×89 for all the evaluated cases. The regions of interest (103 mm×103 mm×80 mm for PG; 79 mm×79 mm×65 mm for SMG) that encompass each PG and SMG were extracted for evaluation. For VTB and VF, because the imaged anatomies were not consistent in PCT and CBCT, we extracted the same anatomic regions in PCT and CBCT for evaluation to ensure anatomy consistency.

Evaluated DIR algorithms

In this work, ten intensity-based deformable image algorithms were evaluated. These DIRs methods were classified into four categories: optical flow based, demons based, level-set based and spline based. Representative algorithms in each category were chosen for validation (listed in Table 2). For the ten selected algorithms, eight (HS [38], HSLK [39], FFD [40], OD [41], MD [42], SFD [43], DFD [44] and LS [45]) were implemented using an open source DIR toolkit-the DIRART [46]. The BSpline method was performed by another open source DIR package, Elastix [47], and the DISC was an in-house GPU-based DIR algorithm developed previously by our group [19].

Grouping and staging

We classified the ROIs into two structural groups, i.e., bony structures including the VTB and the VF, and the soft-tissue structures including the PG and the SMG. Considering that the volume of the PG or the SMG may vary irregularly during the 6~7-week treatment course, we partitioned the entire treatment into six different stages; each stage included five consecutive treatment fractions. ROI volume variations in each stage and its relationship with DIR performance were analyzed.

ROI volume variation

In this study, the right and left parotid or submandibular glands were treated as independent structures (termed as rPG, lPG, rSMG and lSMG, respectively) in light of the fact that they are indeed anatomically isolated (though functioning similarly) and respond separately to incident radiation beams. The relative volume change of each structure in each patient over the entire treatment course was analyzed using linear regression analyses.

Contour propagations

Deformation of ROI contours was performed similar to in Castadot *et al.* [14]. First, all ROI contours on PCT and CBCT were converted to a binary mask where the voxels inside and outside the contours were labeled as one and zero, respectively. After DIR, the resulting deformation vector field (DVF) was used to transform the ROI mask on the PCT and to yield a deformed mask with continuous values between 0 and 1. A threshold of 0.5 was empirically chosen to binarize the deformed mask whose contours were then compared with the ground truths.

Evaluation metrics

To quantitatively evaluate the agreement between the deformed OARs and the reference OARs, three similarity metrics were used: the Dice similarity coefficient (DSC), the percentage error (PE) and the Hausdorff distance (HD). Given the ground truth reference region A , the deformed region B , and their corresponding boundary point sets $\vec{A} = \{a_1, \dots, a_p\}$ and $\vec{B} = \{b_1, \dots, b_q\}$, the DSC is defined as $DSC = 2(A \cap B)/(A + B)$, which ranges from 0 to 1, corresponding to the worst and the best segmentations, respectively. The PE is defined as $PE = (A \cup B - A \cap B)/A$, with 0 representing the best segmentation. The HD is defined as $HD = \max(h(\vec{A}, \vec{B}), h(\vec{B}, \vec{A}))$, where $h(\vec{A}, \vec{B}) = \max_{a \in \vec{A}}(\min_{b \in \vec{B}}(\|a - b\|))$ and $\|\cdot\|$ are the L_2 norms on the points of \vec{A} and \vec{B} .

Statistical analysis

Descriptive statistics were calculated to characterize the volume changes of the PG and the SMG during the treatment course, and linear regression analyses were used to compute the slopes of such changes in the relative volume for each structure in each patient.

Independent samples Kruskal-Wallis (K-W) test was adopted for performance comparisons among all algorithms, for performance variations of the same DIR on four different ROIs and for performance variations of different DIRs on the same ROI. Inter-group comparisons were performed with Mann-Whitney tests. Within each stage, K-W tests and one-way analysis of variance (ANOVA) were conducted for non-parametric and parametric data, respectively, to analyze DIR performance difference.

All statistical analyses were implemented using SPSS 19.0 software (SPSS Inc., Chicago, IL), and the statistical significance level was set at $p = 0.05$. For multiple comparisons, the p-value was adjusted accordingly using the Bonferroni correction method for individual comparison tests in SPSS.

Results

Volume changes

As shown in Fig 1, the volumes of both the PG and the SMG reduced over the course of the treatment, and the left and right parts of the PG or the SMG demonstrated similar volume

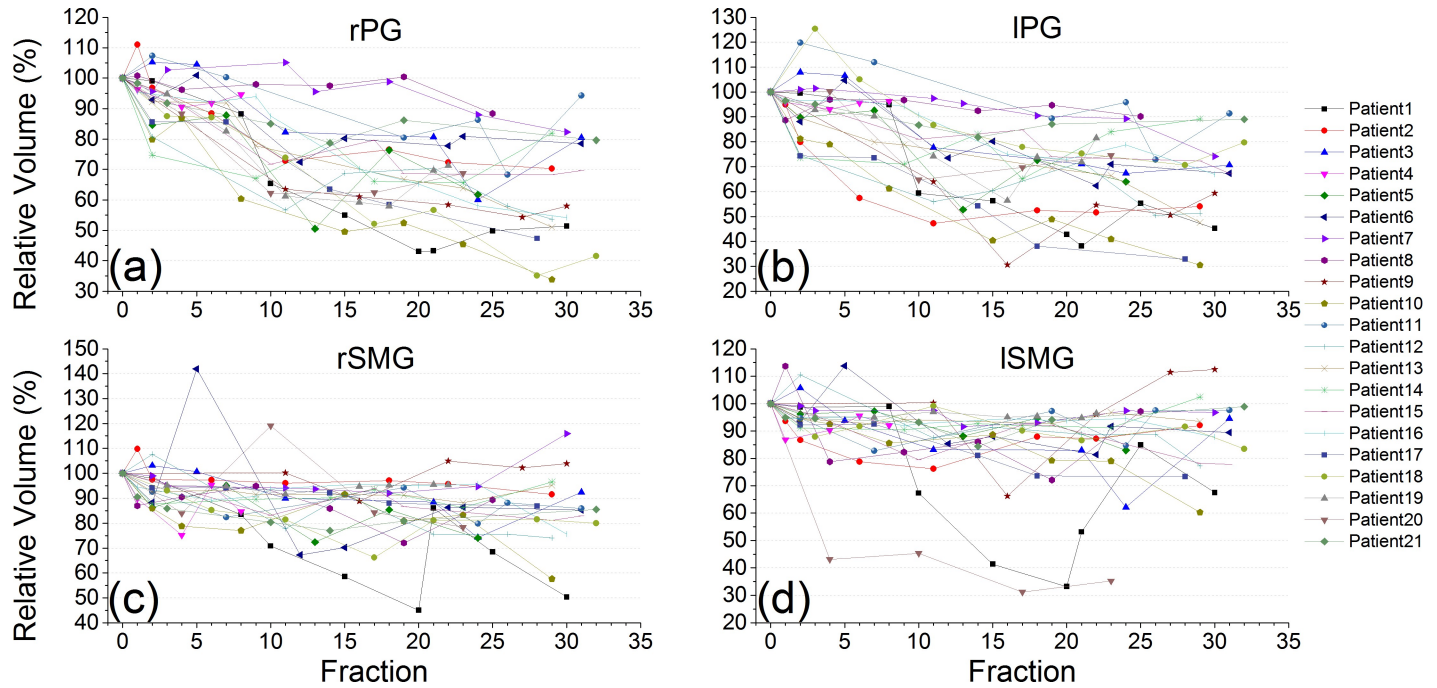


Fig 1. Time trend of the relative volume change for the rPG (a), IPG (b), rSMG (c) and ISMG (d) over the course of the treatment.

<https://doi.org/10.1371/journal.pone.0175906.g001>

reductions. However, the volume of the PG decreased about two times faster than that of the SMG. The mean value, standard deviation (SD), 95% confidence interval (95% CI) and interquartile range (IQR) of the slopes of the relative volume change for each structure are given in Table 3.

Algorithm comparisons

Comparison for all ROIs. Fig 2 shows that comparisons among all algorithms in terms of DSC, PE and HD in all structures. We made the following observations: 1) Rigid registration performed worse than HS, HSLK and FFD ($p < 0.001$, $N = 508$) in terms of DSC and PE. However, rigid registration performed better than OD on all metrics ($p < 0.001$, $N = 508$). 2) OD was the worst DIR algorithm and performed significantly worse than the other DIRs in three metrics ($p < 0.001$, $N = 508$). 3) The optical flow-based algorithms performed best, and the DISC ranked second among all algorithms.

Inter-group and intra-group comparisons. The DIR performance comparison between the bony group and the soft-tissue group is depicted in Fig 3. The Mann-Whitney analysis

Table 3. The mean, standard deviation (SD), 95% confidence interval (95% CI) and interquartile range (IQR) of the slopes of relative volume changes for each structure.

Structure	Mean (%/td*)	SD	95% CI	IQR
rPG	-1.208	0.540	[-1.454;-0.962]	0.898
IPG	-1.170	0.608	[-1.447;-0.894]	0.629
rSMG	-0.546	0.369	[-0.714;-0.378]	0.495
ISMG	-0.471	0.600	[-0.745;-0.198]	0.656

* %/td: percentage per treatment day

<https://doi.org/10.1371/journal.pone.0175906.t003>

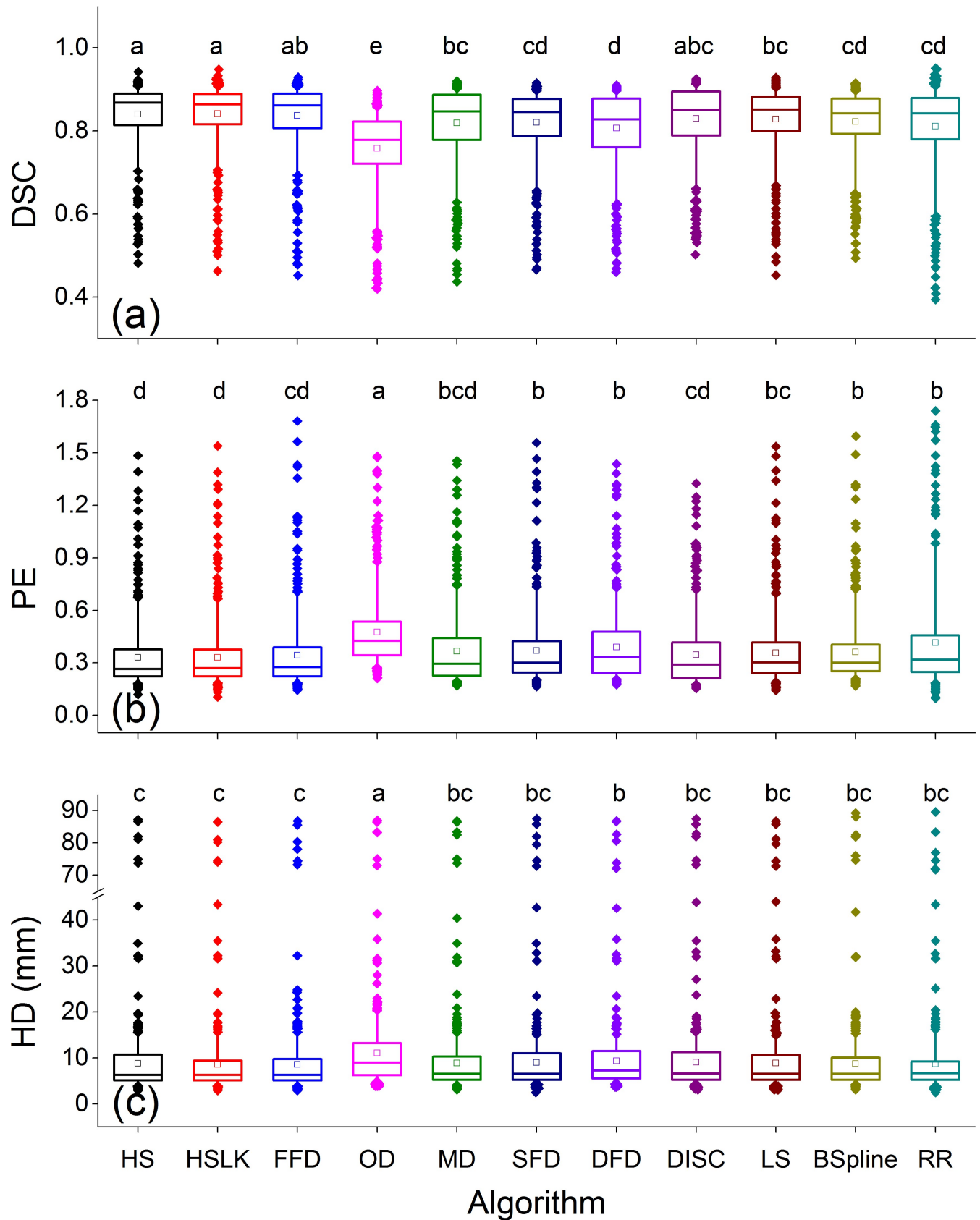


Fig 2. Boxplots of DSC (a), PE (b) and HD (c) by different DIR algorithms for all structures in all treatment stages. The boxes run from the 25th to 75th percentile; the two ends of the whiskers represent the 5% and 95% percentiles of the data, the horizontal line and the square in the box represent the median and mean values, respectively. The diamonds represent outliers. The letters above each box indicate whether a statistically significant difference exists between any two DIRs. No common letter between any two algorithms indicates that the two DIRs are significantly different. RR is an abbreviation for rigid registration.

<https://doi.org/10.1371/journal.pone.0175906.g002>

showed that, for all registration methods, DIR performance in the bony group was better than that in the soft-tissue group in terms of DSC and PE ($p < 0.001$, $N = 254$). In addition, the precision of each algorithm was comparatively high in the bony group where smaller standard deviations of all metrics were found.

In the bony group, all DIRs except for BSpline and OD were better than rigid registration in terms of DSC and PE (DFD: $p = 0.002$ for DSC; LS: $p = 0.019$ for DSC and $p = 0.029$ for PE; SFD: $p = 0.022$ for DSC; others: $p < 0.001$ for DSC and PE, $N = 254$), while in the soft-tissue group, DIRs were comparable to or even worse than the rigid algorithm. DISC and HS/HSLK were the best two DIRs in the bony and soft group, respectively, while the OD performed significantly worse than the other DIR algorithms for all metrics (bony: $p < 0.001$; soft-tissue: MD: $p = 0.005$ for DSC, $p = 0.025$ for PE, $p = 0.013$ for HD; others: $p < 0.001$, $N = 254$).

Comparisons across ROIs. Fig 4 illustrates the performance of the different DIR algorithms on individual ROIs. We made the following observations: 1) For a same DIR, the performance varied significantly among different ROIs ($p < 0.05$, $N = 127$). 2) For the DSC and PE metrics, most DIR algorithms (e.g., DISC ($p < 0.001$), MD ($p < 0.001$) and FFD ($p = 0.007$ and $p = 0.028$) in the VTB for DSC and PE; all DIRs (LS: $p = 0.001$ and $p = 0.006$, others: $p < 0.001$) but OD and BSpline in the VF for DSC and PE; HSLK ($p = 0.004$) in the PG for PE, $N = 127$) were superior to rigid registration in the VTB, VF and PG; however, rigid registration performed better than some DIRs, such as OD, DFD and MD (MD: $p = 0.018$ for HD; others: $p < 0.001$ for all metrics, $N = 127$) in SMG. 3) For each ROI, DIR performance varied significantly across different DIR methods for all metrics. Generally, the optical flow-based algorithms performed comparatively better in the VF, PG and SMG. The demons-based algorithms had a comparatively worse performance in all ROIs, among which OD performed significantly worse than all other algorithms including the rigid method for all assessed metrics ($p < 0.001$, $N = 127$). 4) DISC, a modified demons-based algorithm, was comparable to the optical flow-based methods.

Comparisons across ROIs and stages. Fig 5 depicts the comparison results of all registration methods of each stage for each ROI and the performance variations during the treatment course. We found that 1) For all metrics, the performances of all registration methods in the bony group were more stable than those in the soft-tissue group. For rigid registration, performance in the bony group was comparable with the soft-tissue group in early stages (stage 1 and 2), but better in the subsequent stages. 2) OD performed worst among all methods. 3) Staging comparisons of each registration method varied with ROI.

For bony structures, the performance of all algorithms remained stable over the treatment. In all stages, rigid registration performed worse than some DIR in terms of DSC and PE, such as DISC in the VTB (stage 2: $p = 0.002$ for DSC and PE; stage 4: $p = 0.008$ for DSC and $p = 0.005$ for PE; others: $p < 0.001$) and HS in the VF (stage 1: $p = 0.142$ for DSC and $p = 0.062$ for PE; stage 2: $p = 0.004$ for DSC; stage 4: $p = 0.018$ for DSC and $p = 0.003$ for PE; stage 6: $p = 0.021$ for DSC and $p = 0.008$ for PE; others: $p < 0.001$). Otherwise, the performance of the rigid method was comparable to that of the other DIRs except for OD. In all stages, two demons-based DIRs (DISC and MD) and two optical flow-based DIRs (HS and FFD) performed best in the VTB and in the VF, respectively.

For both the PG and the SMG, the performance of all registration methods on DSC and PE deteriorated as the stage increased, while rigid registration dropped more quickly than DIRs,

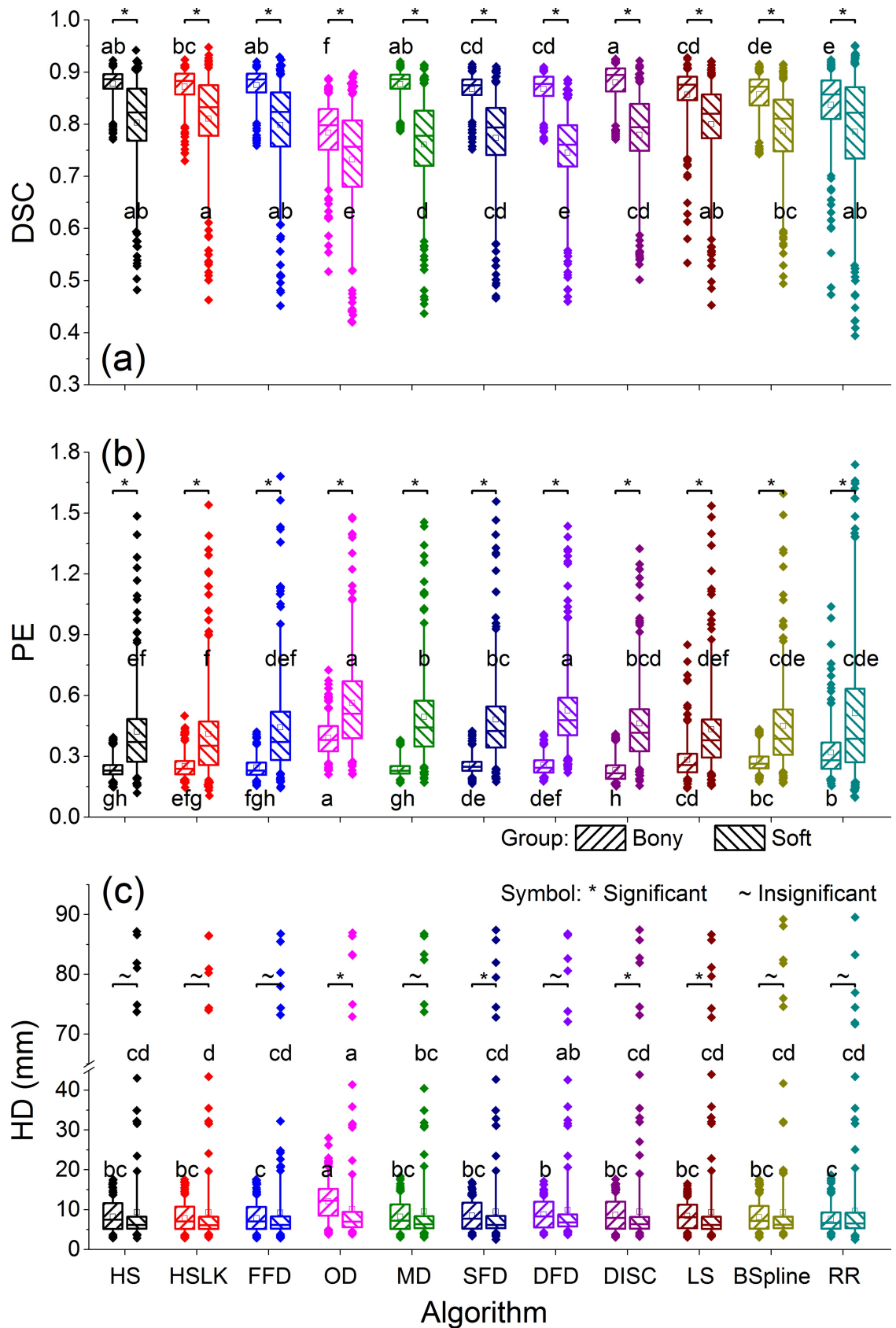


Fig 3. Boxplots of DSC (a), PE (b) and HD (c) by different DIR algorithms for the bony and soft-tissue groups in all treatment stages. The meanings of the symbols in this figure are the same as in Fig 2.

<https://doi.org/10.1371/journal.pone.0175906.g003>

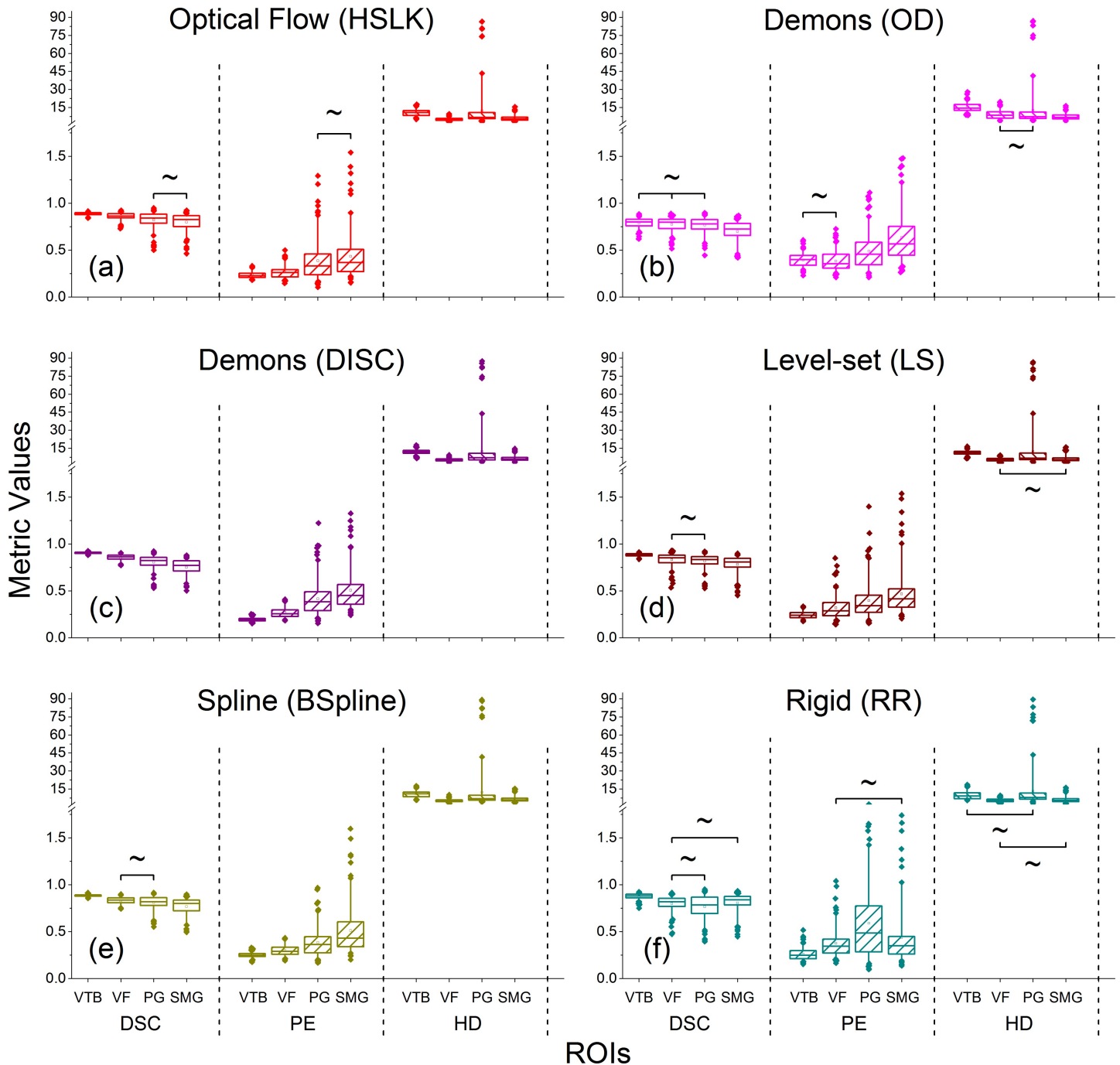


Fig 4. Boxplots of comparisons of performance on different ROIs for representative DIR algorithms, including the HSLK (optical flow-based), OD/DISC (demons-based), LS (level-set-based), BSpline (spline-based) and rigid registration (RR). For each subgraph, no significant difference in pairwise comparisons is marked with “~”, otherwise a statistically significant difference exists. The meanings of each box in this figure are the same as in Fig 2.

<https://doi.org/10.1371/journal.pone.0175906.g004>

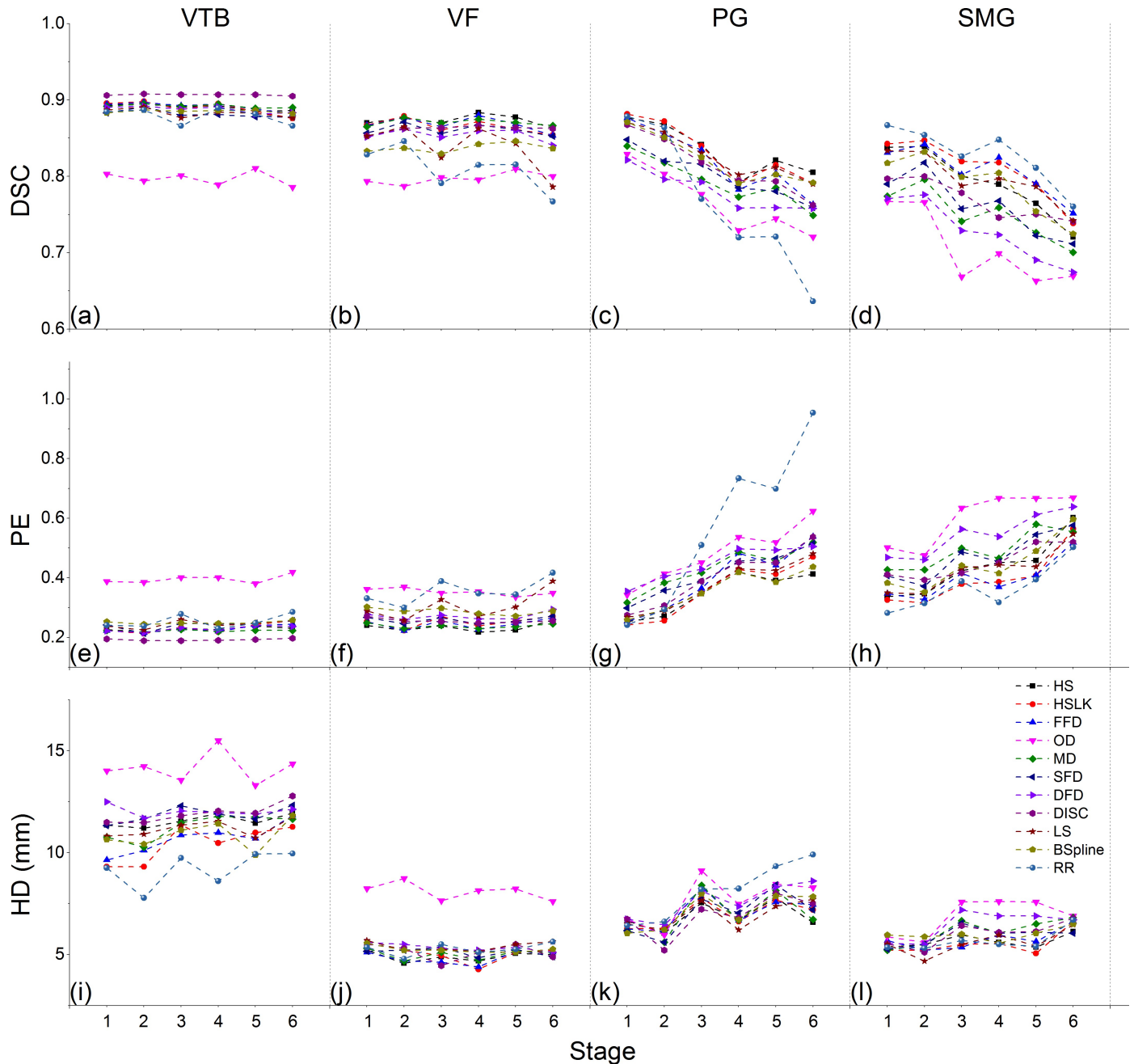


Fig 5. Comparisons across ROIs and stages for different registration algorithms. In each subgraph, the horizontal and vertical axes represent stages ranging from 1 to 6 and the metric values, respectively. Eleven algorithms are marked by different symbols and colors.

<https://doi.org/10.1371/journal.pone.0175906.g005>

especially in the PG. In all stages, the optical flow-based algorithms performed best, followed by the level-set-based or spline-based methods, and the demons-based methods ranked last.

Discussion

In this study, we extensively compared the ability of different DIR strategies for OAR contour propagation between CT and CBCT images. Ten DIR algorithms, including HS, HSLK, FFD,

OD, MD, SFD, DFD, DISC, LS and BSpline, were validated employing planning CT and treatment CBCT images that were retrospectively acquired from twenty-one NPC patients.

Regarding OAR volume change throughout the treatment in NPC patients, we observed similar volume reductions in the PG and the SMG as previously published [4, 25–36]. For instance, Robar *et al.* [25] found that the left and right parotid shrank by 4.7% and 5.0% per week. Comparable results were revealed by Huang *et al.* [35] that the volume of the left and right parotid glands reduced by 7.9%/week and 7.8%/week, respectively. Similarly, Castadot *et al.* [30] reported that the volume reduction of the ipsilateral and contralateral parotid were, respectively, 0.9%/day and 1.0%/day, and 1.5%/day and 1.3%/day for the SMG. In a longitudinal study, Wang *et al.* [29] investigated 82 patients and found that, after receiving three weeks of treatment, the average volume loss at the end of treatment and at two months after treatment were 20.01%, 26.93% and 27.21% for the PG and 11.49%, 16.76% and 16.29% for the SMG, implying that the PG was more sensitive to radiation and would suffer larger volume loss than the SMG.

Before applying the DIR, the treatment table and the thermoplastic mask in all planning CT images and treatment CBCT images were carefully segmented and removed by an experienced physician, leaving only the patient volume in the images, to avoid registration bias induced by such exotic apparatuses. To account for the relationship between DIR performance and volume variation, we compared OARs with changing and constant volumes. Furthermore, we divided the treatment course into different stages for DIR performance analysis. The reasons were threefold, as follows: First, we wanted to determine whether the evaluated DIR algorithms are stable across the entire treatment course, given that some OARs might deform irregularly in different treatment stages. Second, the performance discrepancy between different DIR methods might be related to the different scales of OAR deformation, which might occur in different treatment stages. Third, it would be interesting to know whether DIRs are superior to rigid registration during the entire treatment course.

In many published evaluations regarding DIR performance in H&N patients, DIR is usually believed to be more accurate than rigid registration in OAR contour propagation. For example, Castadot *et al.* [14] compared 12 DIR methods (including the demons-based, level-set-based and fast free-form-based DIRs) by evaluating 18 different ROIs of 5 H&N patients imaged with CT before and during treatment. They found that all DIR strategies corresponded to improved DSC scores when compared with rigid registration ($p < 0.05$, K-W test). Rigaud *et al.* [36] validated 10 different combinations of demons-based and fast free-form-based DIRs with weekly CT images in 15 H&N cancer patients, and they observed that the DSC scores of all DIR methods were remarkably superior to that of rigid registration ($p < 0.01$, ANOVA). Mohamed *et al.* [16] acquired a similar result by utilizing 2720 anatomic ROIs and 50 tumor/nodal ROIs on CT images; they concluded that the 4 assessed DIR strategies (atlas-based, B-spline-based, demons-based and optical flow-based) were all improved over rigid registration ($p < 0.008$ after Bonferroni correction, Steel test).

Interestingly, in contrast to these pilot findings, we found in our study that rigid registration was not always inferior to DIRs (in terms of DSC) as one might expect, especially for low-contrast regions such as the SMG where rigid registration outperformed all the DIRs. This might be attributed to the fact that DIRs are more susceptible in the low-contrast SMG, which tends to fail in providing sufficiently contrasting edges to correctly drive the intensity-based DIRs [48].

For an individual DIR algorithm, the registration performance may vary for different ROIs. For example, the performance of HSLK in the PG was comparable to that in the SMG ($p > 0.05$ for DSC and PE) but worse than that in the VTB ($p < 0.001$ for DSC and PE) and the VF ($p = 0.025$ for DSC and $p < 0.001$ for PE), where HSLK performed better in the VTB than in the

VF ($p = 0.029$ for DSC and $p < 0.001$ for PE). In addition, registration performance between different DIR algorithms also changes with different ROIs. For instance, DISC performed better than HSLK in the VTB but worse in the SMG, while the performance was comparable for the VF and the PG. These findings are in agreement with results of Mohamed *et al.* [16], who found that the differences among the registration accuracies of different structures varies substantially.

The demons-based DIRs (except for the improved demons-based algorithm, the DISC) were found to perform the worst in this study, implying that they might not be suitable for CT-CBCT DIR, although the demons-based DIRs are renowned for their effectiveness and robustness in CT-CT registration [14, 16, 36]. Although CT and CBCT are reconstructed under the same physical principles, the CT-CBCT DIR considered in this study is essentially an inter-modality DIR problem. This poses a great challenge for DIR methods, especially the demons-based DIRs, because of the notorious image quality in CBCT. CBCT suffers more severe image artifacts (e.g., cupping, streak and motion artifacts) and larger scatter contaminations than spiral CT [49–51]. Such artifacts usually result in intensity inconsistency between CT and CBCT, which tends to violate the intensity-consistency assumption in intensity-based demons algorithms.

By staging the treatment course, lower DIR accuracy was found in soft-tissue structures as the treatment proceeded, implying that the DIR performance is influenced by the magnitude of volume loss. This finding coincides with the results of Huger *et al.* [52], who observed evident deteriorating DIR accuracy in as the course of treatment progressed. Moreover, the advantage of DIR over rigid registration was observed in the later stage (stage 3, the 11th–15th fraction) of treatment, when larger volume changes usually occurred. This agrees with previous studies that re-planning should be carried out after 30Gy (i.e., the 15th fraction) [31] or from the 5th or the 15th fraction [35] to account for potential organ deformation.

One limitation of our current work is the exclusion of anatomical landmarks or fiducial markers for geometric accuracy validation. The reasons are twofold: first, all the image data employed were retrospectively collected instead of purposely designed for DIR validation. Thus, all the H&N patients are free of fiducial marker implantation. Second, severe noise and image artifacts in CBCT prevent accurately identifying the anatomic landmarks. Furthermore, landmark migration and stability is a widely debated issue. Distance variations between landmarks within 2 mm (95% confidence) and significant changes in landmark position relative to the tumor over a few weeks were reported [53, 54]. Inaccurate landmark identification will potentially bias the evaluated results.

Conclusions

In this study, we comprehensively compared the ability of OAR contour propagation by ten intensity-based DIR strategies in the context of CT-CBCT DIR using CT/CBCT images retrospectively collected from twenty-one NPC patients. The experimental results indicated that all DIRs evaluated in this work did not necessarily perform better than rigid registration. Generally, the DIR performed better for bony structures than for soft-tissue organs, and the DIR performance tended to vary for different ROIs with different degrees of deformation as the treatment proceeded. Currently, physician-drawn contours are regarded as the ground truths, and human intervention is necessary when a clinically trustworthy DIR algorithm is not yet available. Even DIR algorithms proven to be effective in large-scale clinical data, skipping the contour reviewing procedure remains unlikely in the future to avoid possible failures in rare circumstances. Based on these observations, careful examinations or necessary modifications should be carried out before accepting the auto-propagated contours derived using the DIR algorithms evaluated in this work, especially for automatic re-planning ART systems.

Author Contributions

Conceptualization: XZ LZ.

Data curation: XL YZ YS.

Formal analysis: XL.

Funding acquisition: XZ LZ.

Investigation: XL.

Methodology: XL XZ LZ.

Project administration: XZ LZ.

Resources: YZ YS XZ.

Software: XL YZ YS.

Supervision: XG XZ LZ.

Validation: XL YX.

Visualization: XL SW.

Writing – original draft: XL.

Writing – review & editing: XG XZ LZ.

References

1. Marta GN, Silva V, de Andrade Carvalho H, de Arruda FF, Hanna SA, Gadia R, et al. Intensity-modulated radiation therapy for head and neck cancer: systematic review and meta-analysis. *Radiother Oncol.* 2014; 110: 9–15. <https://doi.org/10.1016/j.radonc.2013.11.010> PMID: 24332675
2. Lee N, Xia P, Fischbein NJ, Akazawa P, Akazawa C, Quivey JM. Intensity-modulated radiation therapy for head-and-neck cancer: the UCSF experience focusing on target volume delineation. *Int J Radiat Oncol Biol Phys.* 2003; 57: 49–60. PMID: 12909215
3. Beadle BM, Liao KP, Elting LS, Buchholz TA, Ang KK, Garden AS, et al. Improved survival using intensity-modulated radiation therapy in head and neck cancers: a SEER-Medicare analysis. *Cancer.* 2014; 120: 702–710. <https://doi.org/10.1002/cncr.28372> PMID: 24421077
4. Barker JL Jr, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys.* 2004; 59: 960–970. <https://doi.org/10.1016/j.ijrobp.2003.12.024> PMID: 15234029
5. Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2006; 64: 355–362. <https://doi.org/10.1016/j.ijrobp.2005.07.957> PMID: 16256277
6. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol.* 1997; 42: 123–132. PMID: 9015813
7. Castadot P, Lee JA, Geets X, Gregoire V. Adaptive radiotherapy of head and neck cancer. *Semin Radiat Oncol.* 2010; 20: 84–93. <https://doi.org/10.1016/j.semradonc.2009.11.002> PMID: 20219546
8. Schwartz DL. Current progress in adaptive radiation therapy for head and neck cancer. *Curr Oncol Rep.* 2012; 14: 139–147. <https://doi.org/10.1007/s11912-012-0221-4> PMID: 22328127
9. Chen AM, Daly ME, Cui J, Mathai M, Benedict S, Purdy JA. Clinical outcomes among patients with head and neck cancer treated by intensity-modulated radiotherapy with and without adaptive replanning. *Head Neck.* 2014; 36: 1541–1546. <https://doi.org/10.1002/hed.23477> PMID: 23996502
10. Lu W, Olivera GH, Chen Q, Chen ML, Ruchala KJ. Automatic re-contouring in 4D radiotherapy. *Phys Med Biol.* 2006; 51: 1077–1099. <https://doi.org/10.1088/0031-9155/51/5/002> PMID: 16481679
11. Thor M, Petersen JB, Bentzen L, Hoyer M, Muren LP. Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol.* 2011; 50: 918–925. <https://doi.org/10.3109/0284186X.2011.577806> PMID: 21767192

12. Janssens G, de Xivry JO, Fekkes S, Dekker A, Macq B, Lambin P, et al. Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy. *Med Phys*. 2009; 36: 4268–4276. <https://doi.org/10.1118/1.3194750> PMID: 19810501
13. Wen N, Glide-Hurst C, Nurushev T, Xing L, Kim J, Zhong H, et al. Evaluation of the deformation and corresponding dosimetric implications in prostate cancer treatment. *Phys Med Biol*. 2012; 57: 5361–5379. <https://doi.org/10.1088/0031-9155/57/17/5361> PMID: 22863976
14. Castadot P, Lee JA, Parraga A, Geets X, Macq B, Gregoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. *Radiother Oncol*. 2008; 89: 1–12. <https://doi.org/10.1016/j.radonc.2008.04.010> PMID: 18501456
15. Hardcastle N, Tome WA, Cannon DM, Brouwer CL, Wittendorp PW, Dogan N, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. *Radiat Oncol*. 2012; 7: 90. <https://doi.org/10.1186/1748-717X-7-90> PMID: 22704464
16. Mohamed AS, Ruangskul MN, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, et al. Quality assurance assessment of diagnostic and radiation therapy-simulation CT image registration for head and neck radiation therapy: anatomic region of interest-based comparison of rigid and deformable algorithms. *Radiology*. 2015; 274: 752–763. <https://doi.org/10.1148/radiol.14132871> PMID: 25380454
17. Xie Y, Chao M, Lee P, Xing L. Feature-based rectal contour propagation from planning CT to cone beam CT. *Med Phys*. 2008; 35: 4450–4459. <https://doi.org/10.1118/1.2975230> PMID: 18975692
18. Nithianathan S, Schafer S, Uneri A, Mirota DJ, Stayman JW, Zbijewski W, et al. Demons deformable registration of CT and cone-beam CT using an iterative intensity matching approach. *Med Phys*. 2011; 38: 1785–1798. <https://doi.org/10.1118/1.3555037> PMID: 21626913
19. Zhen X, Gu X, Yan H, Zhou L, Jia X, Jiang SB. CT to cone-beam CT deformable registration with simultaneous intensity correction. *Phys Med Biol*. 2012; 57: 6807–6826. <https://doi.org/10.1088/0031-9155/57/21/6807> PMID: 23032638
20. Lou Y, Niu T, Jia X, Vela PA, Zhu L, Tannenbaum AR. Joint CT/CBCT deformable registration and CBCT enhancement for cancer radiotherapy. *Med Image Anal*. 2013; 17: 387–400. <https://doi.org/10.1016/j.media.2013.01.005> PMID: 23433756
21. Uneri A, Nithianathan S, Schafer S, Otake Y, Stayman JW, Kleinszig G, et al. Deformable registration of the inflated and deflated lung in cone-beam CT-guided thoracic surgery: initial investigation of a combined model- and image-driven approach. *Med Phys*. 2013; 40: 017501. <https://doi.org/10.1118/1.4767757> PMID: 23298134
22. Zhen X, Yan H, Zhou L, Jia X, Jiang SB. Deformable image registration of CT and truncated cone-beam CT for adaptive radiation therapy. *Phys Med Biol*. 2013; 58: 7979–7993. <https://doi.org/10.1088/0031-9155/58/22/7979> PMID: 24169817
23. Veiga C, McClelland J, Moinuddin S, Lourenco A, Ricketts K, Annkah J, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. *Med Phys*. 2014; 41: 031703. <https://doi.org/10.1118/1.4864240> PMID: 24593707
24. Hardcastle N, van Elmpst W, De Ruyscher D, Bzdusek K, Tome WA. Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy. *Radiat Oncol*. 2013; 8: 243. <https://doi.org/10.1186/1748-717X-8-243> PMID: 24139327
25. Robar JL, Day A, Clancey J, Kelly R, Yewondwossen M, Hollenhorst H, et al. Spatial and dosimetric variability of organs at risk in head-and-neck intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2007; 68: 1121–1130. <https://doi.org/10.1016/j.ijrobp.2007.01.030> PMID: 17398025
26. Han C, Chen YJ, Liu A, Schultheiss TE, Wong JY. Actual dose variation of parotid glands and spinal cord for nasopharyngeal cancer patients during radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008; 70: 1256–1262. <https://doi.org/10.1016/j.ijrobp.2007.10.067> PMID: 18234431
27. Lee C, Langen KM, Lu W, Haimerl J, Schnarr E, Ruchala KJ, et al. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. *Radiother Oncol*. 2008; 89: 81–88. <https://doi.org/10.1016/j.radonc.2008.07.006> PMID: 18707786
28. Vasquez Osorio EM, Hoogeman MS, Al-Mamgani A, Teguh DN, Levendag PC, Heijmen BJ. Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration. *Int J Radiat Oncol Biol Phys*. 2008; 70: 875–882. <https://doi.org/10.1016/j.ijrobp.2007.10.063> PMID: 18262099
29. Wang ZH, Yan C, Zhang ZY, Zhang CP, Hu HS, Kirwan J, et al. Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: a longitudinal study. *Laryngoscope*. 2009; 119: 1966–1974. <https://doi.org/10.1002/lary.20601> PMID: 19688858

30. Castadot P, Geets X, Lee JA, Christian N, Gregoire V. Assessment by a deformable registration method of the volumetric and positional changes of target volumes and organs at risk in pharyngo-laryngeal tumors treated with concomitant chemo-radiation. *Radiother Oncol.* 2010; 95: 209–217. <https://doi.org/10.1016/j.radonc.2010.03.007> PMID: 20385413
31. Cheng HC, Wu VW, Ngan RK, Tang KW, Chan CC, Wong KH, et al. A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patients. *Radiother Oncol.* 2012; 104: 317–323. <https://doi.org/10.1016/j.radonc.2012.03.013> PMID: 22551564
32. Jin X, Hu W, Shang H, Han C, Yi J, Zhou Y, et al. CBCT-based volumetric and dosimetric variation evaluation of volumetric modulated arc radiotherapy in the treatment of nasopharyngeal cancer patients. *Radiat Oncol.* 2013; 8: 279. <https://doi.org/10.1186/1748-717X-8-279> PMID: 24289312
33. Fung WW, Wu VW, Teo PM. Developing an adaptive radiation therapy strategy for nasopharyngeal carcinoma. *J Radiat Res.* 2014; 55: 293–304. <https://doi.org/10.1093/jrr/rrt103> PMID: 23988444
34. Fiorentino A, Cozzolino M, Caivano R, Pedicini P, Oliviero C, Chiumento C, et al. Head and neck intensity modulated radiotherapy parotid glands: time of re-planning. *Radiol Med.* 2014; 119: 201–207. <https://doi.org/10.1007/s11547-013-0326-3> PMID: 24337754
35. Huang H, Lu H, Feng G, Jiang H, Chen J, Cheng J, et al. Determining appropriate timing of adaptive radiation therapy for nasopharyngeal carcinoma during intensity-modulated radiation therapy. *Radiat Oncol.* 2015; 10: 192. <https://doi.org/10.1186/s13014-015-0498-1> PMID: 26377685
36. Rigaud B, Simon A, Castelli J, Gobeli M, Ospina Arango JD, Cazoulat G, et al. Evaluation of deformable image registration methods for dose monitoring in head and neck radiotherapy. *Biomed Res Int.* 2015; 2015: 726268. <https://doi.org/10.1155/2015/726268> PMID: 25759821
37. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* 2012; 30: 1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001> PMID: 22770690
38. Horn BKP, Schunck BG. Determining optical flow. *Artificial Intelligence.* 1981; 17: 185–203.
39. Bruhn A, Weickert J, Schnörr C. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. *International Journal of Computer Vision.* 2005; 61: 211–231.
40. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol.* 2004; 49: 3067–3087. PMID: 15357182
41. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal.* 1998; 2: 243–260. PMID: 9873902
42. Yeo BT, Sabuncu M, Vercauteren T, Ayache N, Fischl B, Golland P. Spherical demons: fast surface registration. *Med Image Comput Comput Assist Interv.* 2008; 11: 745–753. PMID: 18979813
43. Rogelj P, Kovacic S. Symmetric image registration. *Med Image Anal.* 2006; 10: 484–493. <https://doi.org/10.1016/j.media.2005.03.003> PMID: 15896998
44. Wang H, Dong L, O'Daniel J, Mohan R, Garden AS, Ang KK, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol.* 2005; 50: 2887–2905. <https://doi.org/10.1088/0031-9155/50/12/011> PMID: 15930609
45. Vemuri BC, Ye J, Chen Y, Leonard CM, editors. A level-set based approach to image registration. *Mathematical Methods in Biomedical Image Analysis, 2000 Proceedings IEEE Workshop on;* 2000 2000.
46. Yang D, Brame S, El Naqa I, Aditya A, Wu Y, Goddu SM, et al. Technical note: DIRART—A software suite for deformable image registration and adaptive radiotherapy research. *Med Phys.* 2011; 38: 67–77. <https://doi.org/10.1118/1.3521468> PMID: 21361176
47. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging.* 2010; 29: 196–205. <https://doi.org/10.1109/TMI.2009.2035616> PMID: 19923044
48. Yeo UJ, Supple JR, Taylor ML, Smith R, Kron T, Franich RD. Performance of 12 DIR algorithms in low-contrast regions for mass and density conserving deformation. *Med Phys.* 2013; 40: 101701. <https://doi.org/10.1118/1.4819945> PMID: 24089891
49. Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. *Int J Radiat Oncol Biol Phys.* 2007; 68: 522–530. <https://doi.org/10.1016/j.ijrobp.2007.01.038> PMID: 17418960
50. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH. Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. *Med Phys.* 2009; 36: 4755–4764. <https://doi.org/10.1118/1.3223631> PMID: 19928106
51. Hou J, Guerrero M, Chen W, D'Souza WD. Deformable planning CT to cone-beam CT image registration in head-and-neck cancer. *Med Phys.* 2011; 38: 2088–2094. <https://doi.org/10.1118/1.3554647> PMID: 21626941

52. Huger S, Graff P, Harter V, Marchesi V, Royer P, Diaz JC, et al. Evaluation of the Block Matching deformable registration algorithm in the field of head-and-neck adaptive radiotherapy. *Phys Med*. 2014; 30: 301–308. <https://doi.org/10.1016/j.ejmp.2013.09.001> PMID: 24090743
53. Imura M, Yamazaki K, Shirato H, Onimaru R, Fujino M, Shimizu S, et al. Insertion and fixation of fiducial markers for setup and tracking of lung tumors in radiotherapy. *Int J Radiat Oncol Biol Phys*. 2005; 63: 1442–1447. <https://doi.org/10.1016/j.ijrobp.2005.04.024> PMID: 16109463
54. Riboldi M, Orecchia R, Baroni G. Real-time tumour tracking in particle therapy: technological developments and future perspectives. *Lancet Oncol*. 2012; 13: e383–391. [https://doi.org/10.1016/S1470-2045\(12\)70243-7](https://doi.org/10.1016/S1470-2045(12)70243-7) PMID: 22935238