

Article

Evaluating the Quality of Patient-Specific Deformable Image Registration in Adaptive Radiotherapy Using a Digitally Enhanced Head and Neck Phantom

Elisabetta Cagni ^{1,2,*} , Andrea Botti ¹, Matteo Orlandi ¹, Marco Galaverni ³, Cinzia Iotti ³, Mauro Iori ¹ , Geraint Lewis ² and Emiliano Spezi ² 

¹ Medical Physics Unit, Azienda USL-IRCCS di Reggio Emilia, 42122 Reggio Emilia, Italy

² School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

³ Radiotherapy Unit, Azienda USL-IRCCS di Reggio Emilia, 42122 Reggio Emilia, Italy

* Correspondence: elisabetta.cagni@ausl.re.it or cagnie@cardiff.ac.uk; Tel.: +39-0522295987

Featured Application: This paper presents a deformable image registration-based method for the quality assurance of head and neck adaptive radiotherapy using digitally post-processed anthropomorphic phantom image datasets. One of the main findings of this work is that spatial and dose errors are a function of the magnitude of the deformation and of the gradient of the dose distribution. This emphasizes the importance of performing patient-specific deformable image registration verification and, consequently, the need to develop and make available tools that are for this purpose.



Citation: Cagni, E.; Botti, A.; Orlandi, M.; Galaverni, M.; Iotti, C.; Iori, M.; Lewis, G.; Spezi, E. Evaluating the Quality of Patient-Specific Deformable Image Registration in Adaptive Radiotherapy Using a Digitally Enhanced Head and Neck Phantom. *Appl. Sci.* **2022**, *12*, 9493. <https://doi.org/10.3390/app12199493>

Academic Editor: Anna Sarnelli

Received: 1 July 2022

Accepted: 15 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Despite the availability of national and international guidelines, an accurate and efficient, patient-specific, deformable image registration (DIR) validation methodology is not yet established, and several groups have found an incompatibility of the various digital phantoms with the commercial systems. To evaluate the quality of the computed tomography (CT) and on-board cone-beam CT (CBCT) DIRs, a novel methodology was developed and tested on 10 head and neck (HN) patients, using CT and CBCT anthropomorphic HN phantom images, digitally reprocessed to include the common organs at risk. Reference DVFs (refDVFs) were generated from the clinical patient CT-CBCT fused images using an independent registration software. The phantom CT images were artificially deformed, using the refDVFs, and registered with the phantom CBCT images, using the clinical registration software, generating a test DVF (testDVF) dataset. The clinical plans were recalculated on the daily patient 'deformed' CTs, and the dose maps transferred to the patient-planning CT, using both the refDVF and testDVF. The spatial and dosimetric errors were quantified and the DIR performance evaluated using an established operative tolerance level. The method showed the ability to quantify the DIR spatial errors and assess their dose impact at the voxel level and could be applied to patient-specific DIR evaluation during adaptive HN radiotherapy in routine practice.

Keywords: deformable image registration; multimodal image registration; adaptive radiotherapy; head and neck; verification



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

External beam radiotherapy is gradually evolving towards real-time adaptive radiotherapy (ART) [1], which itself is being developed as a new paradigm in radiation oncology [2]. However, ART is, at present, not standardised or widely employed [3]. This is due, not only to the time-consuming processes of the delineation of the targets and organs at risk on the daily computed tomography (CT) images and treatment re-planning, but also to the limited quality of the daily on-board cone beam computed tomography (CBCT) images. Such limitations add practical constraints to the ability to delineate structures manually and estimate the daily delivered dose accurately, and they also contribute to the

limited accuracy of the registration between the daily CBCT and planning CT images [4]. The development of automated tools that enable high-quality ART is likely to lead to significant efficiency gains and foster wider clinical uptake [5].

In the process of delivering automated ART solutions, a challenging task is the validation of the image registration algorithms for clinical use, as their performance depends on the complexity and quality of the images used in the registration task [6]. In the ART workflow, the deformable vector field (DVF), which is generated during the deformable image registration (DIR), could also be applied for dose mapping and dose accumulation purposes. In this process, any uncertainty in the DVF generation would be propagated directly to the calculated dose map.

Considerable research has been done in the investigation of DIR accuracy, primarily through the creation of evaluation metrics. Multi-institutional studies have been performed to validate the use of DIR algorithms [7,8] and these found variations in the DIR performance among different institutions as a result of differences in the procedure, therefore, the accuracy of the DIR was found to be site-specific. A recent survey by Viergever et al. [9] highlighted that the validation of the DIR registration is currently an urgent problem to solve.

The report of the American Association of Physicists in Medicine (AAPM) Task Group 132 (TG132) [10] provided an essential set of guidelines for the quality assurance (QA) and quality control of the image registration operations for the overall clinical process. The TG132 report recommends a series of tests and corresponding metrics that should be evaluated and reported during the image registration software verification.

The TG132 suggests the use of a known DVF to test deformable image registration [10]; this method is generally used with digital phantoms that consist of two image sets linked via a known DVF (ground truth). Several phantom data sets have been generated by the TG132 [10] for use in QA programmes for this purpose. However, a few groups found an incompatibility of certain digital phantoms, provided by the TG132 report, with commercial software [11,12], and/or they found some of the provided phantom images were still limited in their utility. For example, only a pelvic anatomical phantom was made publicly available for DIR verification, and only a single known transformation (refDVF) was provided for test purposes [10]. Pukala et al. [13] reported the differences observed between 10 virtual head and neck (HN) phantoms for DIR verification over five different commercial software systems; these results emphasized the need to assess the DIR accuracy on each considered clinical DIR.

A recent (2021) UK survey [14] highlighted that there is a need for additional guidelines, training and better tools for commissioning DIR software and the evaluation of registration results, which should include the development of recommendations on which quantitative metrics to use. Moreover, in a recent publication, the NRG Oncology medical physics subcommittee working group, composed of nine institutions, reported the limitations of commercial systems in implementing the TG132-recommended metrics [11], due to the inconsistency of the DVF format exported from each individual commercial system with the format of the refDVF provided by the AAPM. The lack of proper Digital Imaging and Communications in Medicine (DICOM) [15] standard definitions for this registration-type limits the direct comparison between the user-provided DVF and the gold-standard refDVF and further limits the interpretability and usability outside the original system.

Shi et al. [16] investigated the correlation between contour-based metrics and DVF errors. They reported a moderate correlation between the contour propagation and accurate dose deformation, and the summation/accumulation within each contour. Therefore, for adaptive radiation therapy, the accurate dose deformation and summation/accumulation within each contour is determined by the DVF accuracy.

Several studies have been conducted to validate the reliability of DVF accuracy. Veiga et al. [17,18] tested the transformations that mapped the anatomy between the planning CT and CBCT, which were obtained using four different DIR approaches. The data from five HN patients were used to evaluate the performance of each implementation, based on

geometrical matching, the physical properties of the DVFs and the similarity between the warped dose distributions. They found a variability between the different DIR algorithms used, in terms of impact on the uncertainties in the dose warping.

As reported in Paganelli et al. [19], the main limitation of these approaches relies on the fact that a deformation field may fulfil the requirements of physical fidelity, but still differ from the underlying ground truth of the deformation, without conveying any information on the DIR uncertainties and related dose errors. Indeed, a global validation of user-obtained DVF against ground-truth DVF is considered the most accurate DIR evaluation method, yet the ground-truth DVF is not always available in a clinical setting.

This paper is focused on DIR verification for HN adaptive radiotherapy. Practical guidelines and dedicated phantoms for clinical implementation are still needed to test the accuracy of deformable image registration (DIR) in various clinical situations. The registration between the CT and CBCT image sets can be considered a multimodal registration problem. The consistency of the pixel intensity values between the two modalities is restricted because the image quality of CBCT is poorer than that of CT [20], and CBCT images are commonly affected by artefacts due to scatter contamination and truncated CBCT patient volumes [21].

The use of digital phantoms for CT-CBCT DIR QA is limited by the fact that the applied transformations are not patient-specific but are based on pre-defined DVFs. Furthermore, the deformed images of digital phantoms do not have any representative noise variation, inherent in the CBCT images of an actual patient [6]. It is possible to overcome such limitations using images of a physical phantom acquired with the systems and acquisition protocols used in clinical practice. Physical phantoms do not reflect the complexity of real human anatomy, which is affected by many factors, such as the setup variation, organ motion and organ deformation. These factors contribute to the inherent differences between CT and CBCT images, even if both image sets are obtained on the same day. However, the use of a rigid anthropomorphic phantom allows the acquisition of image datasets with identical anatomy for both the CT and CBCT scans. This is the main reason for the use of physical anthropomorphic phantom images in the presented method instead of real patient images: the patient's CT and CBCT scans are always different, due to patient organ motion and setup uncertainties. This ensures robust and reliable comparisons by providing a reference, as any discrepancies between the primary CT and CBCT images would be purely related to the performance of the DIR algorithm.

In this study, the images of a commercial anthropomorphic phantom to include common OARs as recommended by international guidelines [22] were processed. The aim of this study was to build and test a novel workflow for the verification of patient-based DIR in HN ART, using the developed phantom images.

2. Materials and Methods

2.1. Digitally Enhanced Phantom

In this study, the ATOM Max Dental and Diagnostic Head Phantom Model-711 (CIRS, Norfolk, VA, USA) (HN phantom) was used. The HN phantom approximates to the average male human head in both size and structure. The original phantom includes detailed 3D anthropomorphic anatomy, such as the brain, bone, larynx, trachea, sinus, nasal cavities and teeth. The bones contain both cortical and trabecular components. The teeth include the distinct dentine, enamel and root structures, including the nerves. The sinus cavities are fully open.

The HN phantom was scanned with both CT and CBCT. The CT image datasets were acquired using a GE Hi-Speed (General Electric Company, Boston, MA, USA) scanner, with the acquisition parameters set in agreement with our clinical HN protocol (120 KV, 100 mA, 2 mm slice thickness). As a second step, the CBCT image sets were acquired on a TrueBeam STx linear accelerator (Varian Medical Systems, Palo Alto, CA, USA), using our HN image acquisition protocols (HN modality). To acquire the entire head of the phantom, a multi-scan CBCT was performed, and the data were combined using the TrueBeam STx advanced

reconstruction module. For the CBCT images, the linear attenuation to the Hounsfield Unit (HU) values calibration was done using scans of the Catphan 504 phantom, following the Varian TrueBeam acceptance procedure manual [23], by assigning the HU values from the Catphan manual to the reconstructed raw image values of the seven material inserts, followed by a linear fit.

Both the CT and CBCT image sets were post-processed to digitally insert several OARs which are not present in the original phantom, including the brainstem, oral cavity, left and right parotid glands, larynx and eyes. Firstly, a radiation oncologist (MG) drew outlines of these anatomical structures on the phantom CT and CBCT images, following international guidelines [22], for their standard shape, anatomical position and volume.

Subsequently, the original HU values were replaced with surrogate HU values for all the above-mentioned structures, using an in-house developed script based on the MATLAB code (The Mathworks, Natick, MA, USA) and the Computational Environment for Radiological Research tool (CERR, <http://www.github.com/cerr/CERR>, accessed on 20 February 2020) [24]. The surrogate HU value in each voxel, i , of a given organ OAR, $HU_{processed\ OAR_i}^{phantom}$, was derived using the following equation:

$$HU_{processed\ OAR_i}^{phantom} = \left(\overline{HU_{OAR}^{pt}} - \overline{HU_i^{pt}} \right) + HU_{original\ OAR_i}^{phantom} \quad (1)$$

where $\overline{HU_{OAR}^{pt}}$ and $\overline{HU_i^{pt}}$ are the mean HU values over the voxels included in a ROI belonging to a specific organ ($\overline{HU_{OAR}^{pt}}$) and the tissue around that organ ($\overline{HU_i^{pt}}$), averaged over a set of 10 real patient CT or CBCT images, respectively; $HU_{original\ OAR_i}^{phantom}$ is the HU value in the original HN phantom CT or CBCT image for each voxel.

The $\overline{HU_{OAR}^{pt}}$ and $\overline{HU_i^{pt}}$ parameters were measured by taking the mean value of several (typically four) square ROIs of 10×10 pixels, placed in different positions inside the OAR and its surrounding tissue, to reflect the typical HU variation in real patient images. These mean values were then averaged over a set of 10 pairs of actual patient CT and CBCT images. This process was done separately for the CT and CBCT image sets (i.e., for each patient, one CT and one CBCT image were considered). Using this method, the intrinsic noise of each image type, the CT and CBCT, was maintained in the digitally modified phantom image sets. The post-processed phantom images are in the standard DICOM [15] format, in order to be compatible with different image registration systems. In the rest of the manuscript, the phantom image set is referred to as the post-processed enhanced phantom CT.

2.2. Workflow

The method used to validate the DIR applied to the CT-CBCT image pairs was based on a known transformation for each voxel [6]. This method requires an independent ‘third-party’ image registration software system to generate the reference DVF. We will refer to the third-party software as the ‘reference DIR software’, as it will be used to generate a reference DVF. The only requirement for the reference and the clinical DIR software systems should be the independence of each other in the DIR procedure to avoid biasing the results. As a guarantee of the correctness of the procedure, it is not essential that the reference software should be any more accurate in the DIR process than the clinical one.

The set of selected clinical patient images (CT and CBCT) that must be tested are registered using the reference software (see Figure 1a, spatial analysis workflow). This registration is considered as the reference for the patient CT-CBCT data sets (refDVF). As a second step, the HN phantom CT images are artificially deformed into ‘warped’ phantom CT (wCT) images by applying the refDVF transformation. Using the clinical DIR software, the phantom CBCT image is registered with the wCT phantom image, generating a test DVF to be evaluated (testDVF). Note that the term refDVF relates to the patient image

registration throughout, whereas the term testDVF always refers to the warped phantom CT and phantom CBCT DIR.

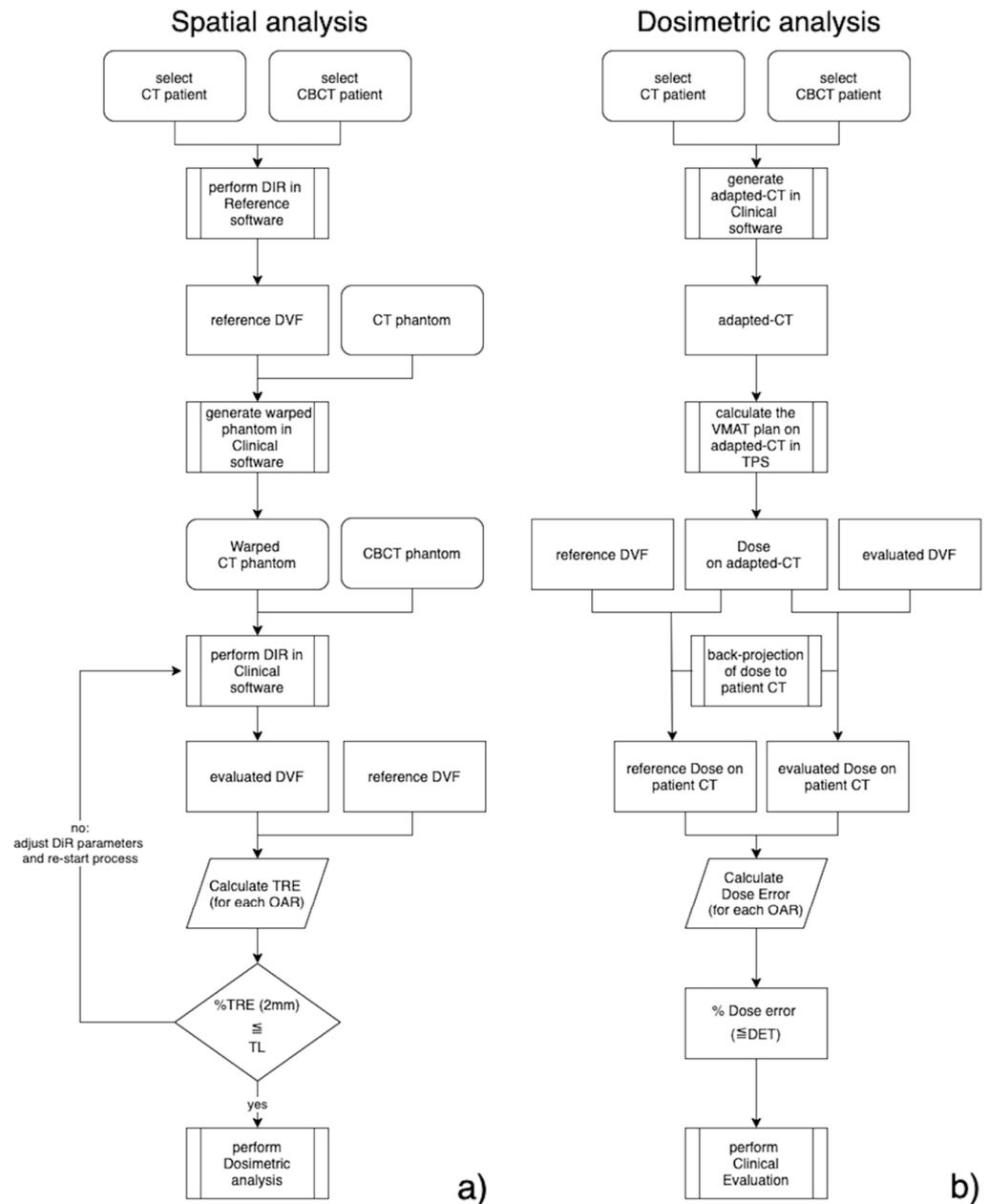


Figure 1. Schematic workflow of CT-CBCT DIR validation method for spatial and dosimetric analysis. **(a)** Spatial Analysis. The HN patient CBCT-CT DIR was considered as reference DVF (refDVF). The refDVF was generating using third-party reference software. Post-processed phantom CT images were digitally warped using the reference DVF resulting in warped phantom CT scans. The phantom CBCT and warped CT images were then registered in the clinically used DIR software, generating an evaluated DVF (testDVF). The testDVF was compared with the refDVF by using a TRE metric and an operative tolerance level (OTL) analysis performed to quantify the accuracy of the DIR. **(b)** Dosimetric Analysis. Treatment plans were calculated in the TPS on each patient’s adapted-CT image (patient CT image warped into CBCT image) resulting in a delivered dose. The delivered dose was transferred back to the patient’s planning CT image via both the refDVF (reference dose) and the testDVF (test dose). The dose difference between the two was calculated with a clinical assessment, involving a radiation oncologist, required for a final DIR validation.

The quality of the registration was assessed as the ability of the clinical software to replicate the refDVF, artificially introduced into the phantom wCT images. The target registration error (TRE) was computed as a metric to evaluate the difference between the refDVF and the testDVF (Section 2.5). As recommended by the TG132, the percentage of voxels with a TRE within 2 mm was compared with a designated Operative Tolerance Level (OTL) value (Section 2.6). If the testDVF does not overcome the OTL analysis, the evaluated registration process should be re-started, adjusting certain registration parameters, such as the volume of interest.

The propagation of dosimetric errors due to differences in the DIR maps is schematized in Figure 1b, labelled as the dosimetric analysis workflow. The VMAT clinical treatment plans were recomputed in the clinical TPS for each patient's CBCT adapted-CT images to avoid bias in the dose calculation (see Section 2.4), and back-projected to the corresponding patient's CT images, via both the refDVF and testDVF, generating reference dose (refDose) and evaluated dose (testDose) arrays, respectively. The dosimetric differences between the refDose and testDose should be evaluated in conjunction with clinical radiation oncologists.

2.3. Clinical Data

Ten clinical patients with advanced oropharyngeal cancer who underwent radiotherapy at our institution were randomly selected from the departmental database. All the plans used volumetric modulated arc therapy (VMAT, RapidArc) to treat three targets at dose levels of 69.96, 59.40 and 54.45 Gy in 33 fractions, with a simultaneous integrated boost generated with an Eclipse Treatment Planning System (TPS) v.13.6 (Varian Medical Systems, Palo Alto, CA, USA). Bilateral neck irradiation was delivered to all the patients, with involved high-dose PTV on the right side for six patients, the left side for two patients, and centrally for the remaining two patients. For each treatment fraction, a CBCT scan was acquired, using the clinical HN protocol, before delivering treatment to assess the patient setup and anatomical variation. All the CBCT images were automatically saved in an ARIA database (Varian Medical Systems, Palo Alto, CA, USA). The published studies have shown that the anatomic changes in HN cancer patients are more pronounced in the first half of treatment [25] and, based on our clinical institute experience, that re-planning requests are often performed in the second half of treatment. Hence, for each patient, the CBCT scans performed around the middle (fraction #16) and latter parts of the treatment course (fraction #26) were extracted from the imaging database for further analysis.

2.4. Demonstration of Clinical Application

The extracted pairs of CBCT and planning CT images for the 10 clinical patients selected were imported into MICE Toolkit v.1.1.0 (NONPI Medical AB, Stockholm, Sweden), which was used in this study as the reference software. The MICE Toolkit was used to produce the refDVF, being the only independent DIR software available in our institute for comparison with our clinical system (Velocity AI, described below). Within the MICE, the open-source Elastix software module [26] was employed. The registration method used in this case is of the same type as the one in the clinical platform but implemented independently. It consists of a B-spline method with an adaptive stochastic gradient descent optimiser, with the number of resolutions equal to 3, the maximum number of iterations of 1000 and an interpolator order of 3, similar to those parameters reported in the literature [27]. The DVF from each DIR (in total, 20 DVFs) was subsequently exported in the MATLAB format, converted to the binary format, and used as the refDVF.

The Velocity AI Oncology Imaging Informatics System v.4.0 (Varian Medical System, Palo Alto, CA, USA) was the DIR module used in our institute in clinical practice. The post-processed phantom scans were imported into the Velocity AI and linked to each of the 10 patients selected for the study. Both the phantom CT and CBCT images were rigidly registered to the CT image of the patient. This was done to align the phantom CT image with the clinical refDVF to produce a more clinically realistic warped phantom CT image.

Following the workflow described in Figure 1, using the Velocity AI, the HN phantom CT images were artificially deformed into wCT images, using each of the imported refDVF. Then, the phantom CBCT image was registered with the phantom wCT image, generating a testDVF obtained using the Velocity AI algorithm. The ‘CBCT corrected multi-pass deformable’ modality in the Velocity AI was used to generate the testDVF. This applies a ‘fade correction’ prior to the registration, which enhances the low signal regions in the CBCT image [28]. The Velocity AI’s primary registration algorithm uses a multi-resolution approach based on mutual information. The transform is a cubic B-spline, the interpolator is a bi-linear interpolation function, and the optimiser is based on the steepest gradient descent method [28]. The testDVF was then exported from the Velocity AI in a binary format. This workflow was repeated for all 20 registrations for the 10 patients. To ensure that the simulated deformation could be used as a reference to validate the image registration algorithm, for each case the warped phantom CT images were carefully inspected to verify that the changes induced in the phantom scans produced by the refDVF were realistic. Each testDVF was compared to the refDVF using the MATLAB in-house scripts. The quality of the registration was assessed as the ability of the testDVF to replicate the artificially generated refDVF.

For each patient and treatment fraction considered, an adapted-CT image of each refDVF was generated in the Velocity AI. The adapted-CT images used the HU values of the patient CT image, mapped onto the CBCT image, to avoid bias in the dose calculation [28]. The adapted-CT images were then imported into the Eclipse TPS, and the VMAT clinical treatment plans were recomputed on each patient’s adapted-CT images and sent back to the Velocity AI. To assess the dose propagation errors due to the differences in the DIR maps, the dose on the adapted-CT was back-projected to the corresponding patient CT images via both the refDVF and testDVF.

2.5. Data and Statistical Analysis

The spatial and dosimetric errors were calculated for all the voxels contained within the following OARs: the brainstem, spinal cord, mandible, left parotid, right parotid, larynx and oral cavity, as well as the external body contours. In two of our ten patients, the larynx was missing as an OAR because those patients were affected by laryngeal tumours, which were included in the clinical target volumes.

The spatial errors were calculated using the TRE [6], which describes the difference between the co-located voxels once they have been transferred through the refDVF and testDVF. The TRE values were evaluated as a function of the treatment fraction number and the refDVF magnitude on a voxel basis. The percentage of voxels with a TRE of less than or equal to 2 mm was considered in the analysis as a significant parameter for the spatial errors, as was suggested by the AAPM-TG132 [10].

Dose errors were calculated on the patient CT images as being the difference between the back-projected doses, using the two propagation methods (the reference vs. the evaluated). The dose errors were evaluated on a voxel basis. As a significant metric for the dose errors, the percentage of voxels within the dose error of a specific dose error threshold (DET) was calculated. In this study, a significant value for the DET was considered as 5% of the prescribed dose of 70 Gy, thus corresponding to 3.5 Gy.

The Pearson’s correlation coefficient metric was used to assess the statistical significance of the correlation of the observed TRE, as a function of the refDVF magnitude. The correlation between the dose error, TRE, dose gradients and the combination (scalar product) of these two (the TRE and the dose error) was also assessed using the Pearson’s correlation, as in a previously published similar study [17]; p -values < 0.05 were considered significant.

2.6. Tolerance Level Based on AAPM-TG132

The AAPM TG132 Report [10] provided digital phantoms for use in commissioning and quality assurance programmes for image registration accuracy tests. For rigid registra-

tion, TG-132 proposed 2 data sets created using the ImSimQA (Oncology Systems Limited, Shrewsbury, UK) (basic phantom data set and anatomic data set (a pelvic phantom)) for various modalities (CT, CBCT, PET, MRI-T1, and MRI-T2) for rigid registration. For the DIR accuracy assessment, the TG-132 provides a dataset pair of anatomical pelvic phantom images comprising:

1. a basic anatomical dataset (CT),
2. a basic deformation dataset, same as basic anatomical dataset with added gaussian noise variation to simulate CBCT image,
3. a ground-truth DVF transformation (TG132-refDVF) file in the DICOM format.

The basic anatomical phantom CT provided by the AAPM was deformed by using the provided TG132-refDVF. The registration between the warped anatomical phantom CT and basic deformed phantom (CBCT) was performed in the Velocity AI, generating the TG132-testDVF. Differences between the TG132-refDVF and TG132-testDVF were quantified in terms of the target registration error (TRE). The results of this test were used to establish an OTL for the DIR accuracy in the clinical dataset used in this study.

3. Results

3.1. Digitally Enhanced Phantom

The original and post-processed CT and CBCT phantom images in three different transverse sections are shown in Figure 2. Table 1 reports the HU values of OARs in the original and processed CT and CBCT phantom images. Both post-processed phantom image sets have been made publicly available as additional downloadable material for this paper, as Supplementary Material.

Table 1. Original and post-processed CT and CBCT transverse sections of the CIRS ATOM Max Dental and Diagnostic Head Phantom Model-711 used in the method presented for QA of DIR. The HU values of parotid glands, larynx, oral cavity, brainstem and eyes were digitally modified using MATLAB and CERR tools, as described in Section 2.1. HU values for original and post-processed CT and CBCT phantom images reported in the table referred to a ROI of 20×20 pixels.

| ORGAN | Original Phantom CT Images (HU) | | | | Processed Phantom CT Images (HU) | | | |
|----------------|-----------------------------------|------|-------|-------|------------------------------------|------|-------|-------|
| | Mean | Std | Min | max | mean | Std | min | Max |
| Parotid Glands | 50.0 | 6.0 | 35.0 | 73.0 | 26.4 | 7.4 | −1.0 | 53.0 |
| Larynx | 42.2 | 5.8 | 25.0 | 57.0 | 72.5 | 28.5 | 21.0 | 108.0 |
| OralCavity | 48.1 | 4.2 | 37.0 | 60.0 | 40.0 | 5.3 | 25.0 | 53.0 |
| BrainStem | 48.4 | 7.7 | 32.0 | 74.0 | 38.6 | 13.5 | 10.0 | 66.0 |
| Eyes | 75.1 | 8.4 | 53.0 | 100.0 | 101.9 | 9.5 | 62.0 | 135.0 |
| ORGAN | Original Phantom CBCT Images (HU) | | | | Processed Phantom CBCT Images (HU) | | | |
| | Mean | Std | Min | max | mean | Std | min | Max |
| Parotid Glands | 61.6 | 14.8 | 17.0 | 100.0 | 18.1 | 14.1 | −16.0 | 56.0 |
| Larynx | 112.9 | 14.8 | 58.0 | 154.0 | 185.9 | 66.6 | 85.0 | 278.0 |
| OralCavity | 88.3 | 19.1 | 25.0 | 147.0 | 96.7 | 18.8 | 39.0 | 150.0 |
| BrainStem | 91.7 | 27.7 | 23.0 | 182.0 | 44.8 | 65.2 | −85.0 | 182.0 |
| Eyes | 27.7 | 20.5 | −28.0 | 92.0 | 40.9 | 20.5 | −16.0 | 104.0 |

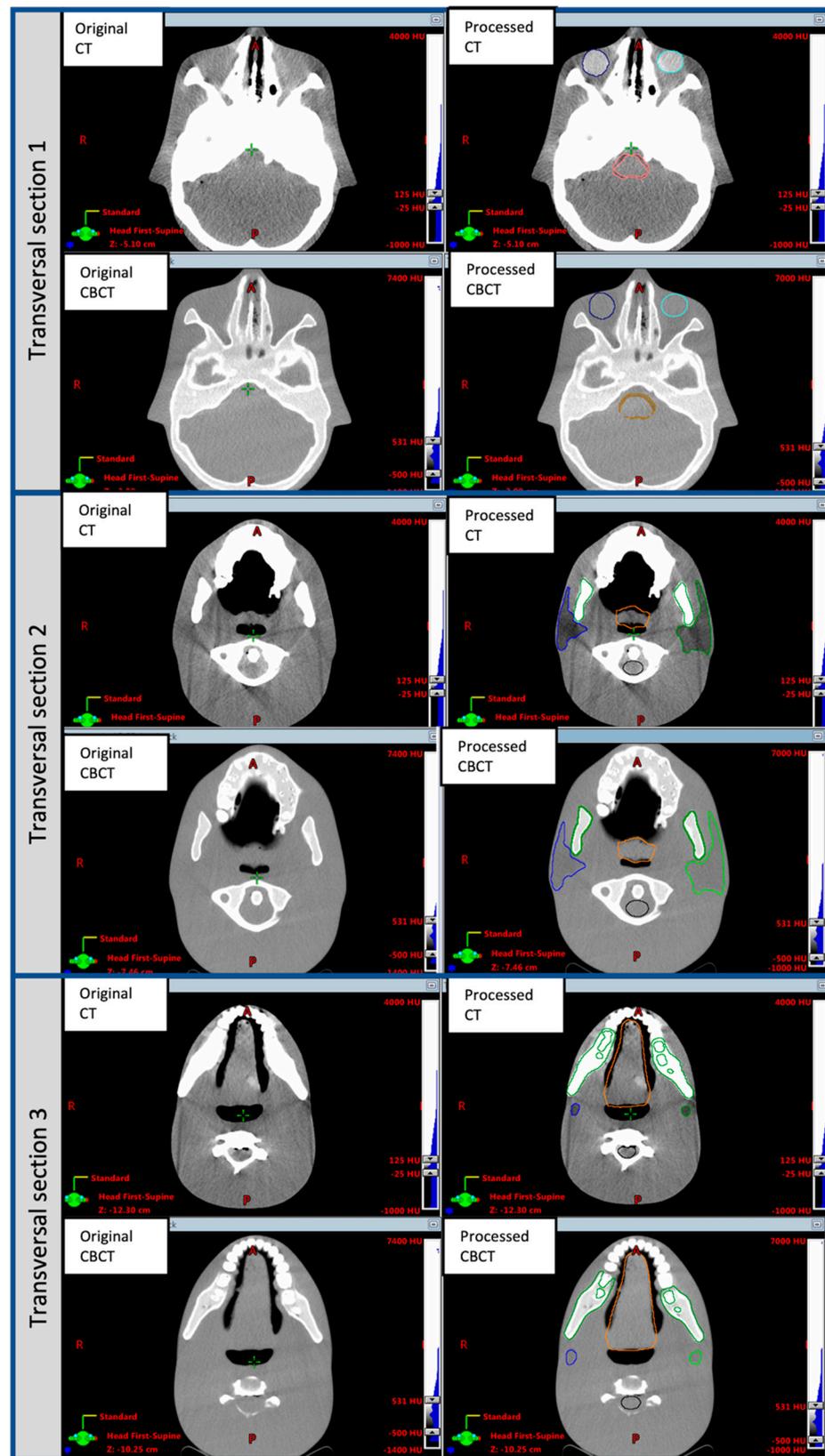


Figure 2. Original and post-processed CT and CBCT transverse sections of the CIRS ATOM Max Dental and Diagnostic Head Phantom Model-711 used in the method presented for DIR accuracy evaluation. Sections were chosen to visualize organs of interest. The HU values of parotid glands, larynx, oral cavity, brainstem and eyes were digitally modified using MATLAB and CERR tools, as described in Section 2.1.

3.2. Reference Deformable Vector Fields

A visual inspection of the deformed patient images and DVF magnitude was performed after each registration using the designated reference (MICE Toolkit) software to check the algorithm performance and the integrity of the resulting DVF. The magnitudes of the refDVF (mean \pm SD) over all fractions and patients were: 7.56 ± 2.33 mm, 4.44 ± 3.26 mm, 5.17 ± 4.94 mm, 3.67 ± 1.5 mm, 3.70 ± 1.33 mm, 3.77 ± 2.44 mm, 12.57 ± 8.94 mm and 5.52 ± 2.54 mm for the body, left parotid gland, right parotid gland, larynx, oral cavity, mandible, brainstem and spinal cord, respectively.

3.3. Application of the Method: Spatial and Dosimetric Analysis

The TRE values for each OAR and each registration are plotted in Figure 3a. The TRE was found to vary across the OARs of interest. The mean and standard deviation values (mean \pm SD) of the TRE over all fractions and patients were 4.6 ± 4.6 mm, 2.6 ± 1.4 mm, 2.7 ± 1.5 mm, 2.1 ± 1.0 mm, 2.0 ± 1.2 mm, 1.8 ± 1.2 mm, 5.6 ± 3.3 mm and 2.5 ± 2.0 mm for the body, left parotid gland, right parotid gland, larynx, oral cavity, mandible, brainstem and spinal cord, respectively. The large range of values of the TRE for the brainstem were due to the border of the ROI exceeding the CBCT boundary in the cranial direction in the MICE toolkit, when any mask was used. This could have induced unnatural and drastic shearing of the refDVF for the brainstem in the cranial section of the CBCT boundary, as reported in the refDVF data in Section 3.2. The testDVF generated with the Velocity AI showed a smoother DVF in that area, resulting in consequently larger TRE values. The differences in the TREs between the two fractions, compared on a voxel basis, were found to be insignificant using the Wilcoxon signed-rank test, with a p -value ≈ 1 .

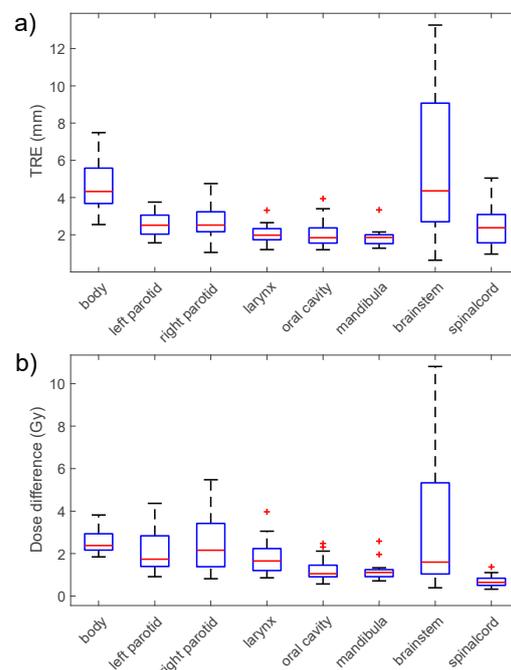


Figure 3. Boxplots of (a) TRE difference and (b) dose difference distribution sorted by OAR. For each box, the central mark represents the median value, while the bottom and top edges of the box are the 25th and 75th percentiles, respectively. The ‘whiskers’ represent the range of values. Observations beyond the whisker length are marked as outliers (+). By definition, an outlier is a value that is more than 1.5 times the interquartile range away from the bottom or top of the box. An outlier appears as a red + sign.

The dose error values for each OAR are plotted in Figure 3b. Similar to the TRE analysis, the median dose error value varied according to the OAR of interest. The mean \pm SD dose difference values over all fractions and patients were 2.0 ± 4.3 Gy, 2.2 ± 2.4 Gy,

2.5 ± 2.7 Gy, 1.8 ± 1.8 Gy, 1.3 ± 1.6 Gy, 1.2 ± 1.8 Gy, 3.2 ± 2.6 Gy, 0.7 ± 0.7 Gy for the body, left parotid gland, right parotid gland, larynx, mandibula, brainstem and spinal cord, respectively.

3.4. Baseline AAPM-132 Tolerance Levels Analysis

The mean TRE errors (mean \pm SD) between the TG132-refDVF, provided by the Report, and the TG132-testDVF, generated by the Velocity AI, were 2.31 ± 1.13 mm for the phantom external contour (body). The overall % of voxels with a TRE of less than or equal to 2 mm was 45% (rounded to nearest 1%). This percentage value was considered as the OTL for the registrations evaluated in this study. Since the AAPM Report did not provide any dose error tolerance level information, the OTL could only be considered in terms of the spatial accuracy.

Figure 4 reports, in the correlation graph below, the percentage of voxels with a TRE error ≤ 2 mm versus the percentage of dose errors \leq the DET (5% of prescribed dose in this analysis). The Figure 4 x-axes show the percentage of voxels with a TRE within 2 mm (as suggested by the AAPM-TG132) across the 20 registrations for each OAR. Significant variability was observed for all the organs and registrations. The markers '+' indicate the cases for which the percentage of voxels with a TRE < the 2 mm OTL, represented as a vertical red line in the graph. The percentages higher than the OTL (indicated as a 'circle' marker) were found in 30% (body), 55% (left parotid), 40% (right parotid), 31% (larynx), 85% (oral cavity), 100% (mandibula), 25% (brainstem) and 80% (spinal cord) of all the cases (N = 20), for all registrations. The Figure 4 y-axes report the corresponding dose error percentage within 3.5 Gy (5% of prescribed dose), across every registration for each OAR. The percentages of $66.3 \pm 5.8\%$, $63.9 \pm 16.1\%$, $62.0 \pm 17.5\%$, $69.3 \pm 16.4\%$, $82.2 \pm 11.1\%$, $85.1 \pm 8.0\%$, $66.7 \pm 28.2\%$ and $93.6 \pm 6.2\%$ were found for the body, parotids left and right, larynx, oral cavity, mandible, brainstem and spinal cord, respectively. Overall, the correlation graphs showed a generally higher proportion of voxels involving a dose error of less than 5% rather than a TRE of less than 2 mm (visualised as markers above the diagonal bisector line of each panel). The relation between the TRE and dose error percentage is quite complex and depends on a single registration (see also Section 3.5). For almost all OARs, some registrations passed the spatial OTL analysis but showed a quite low dose error <5% percentage, and vice versa.

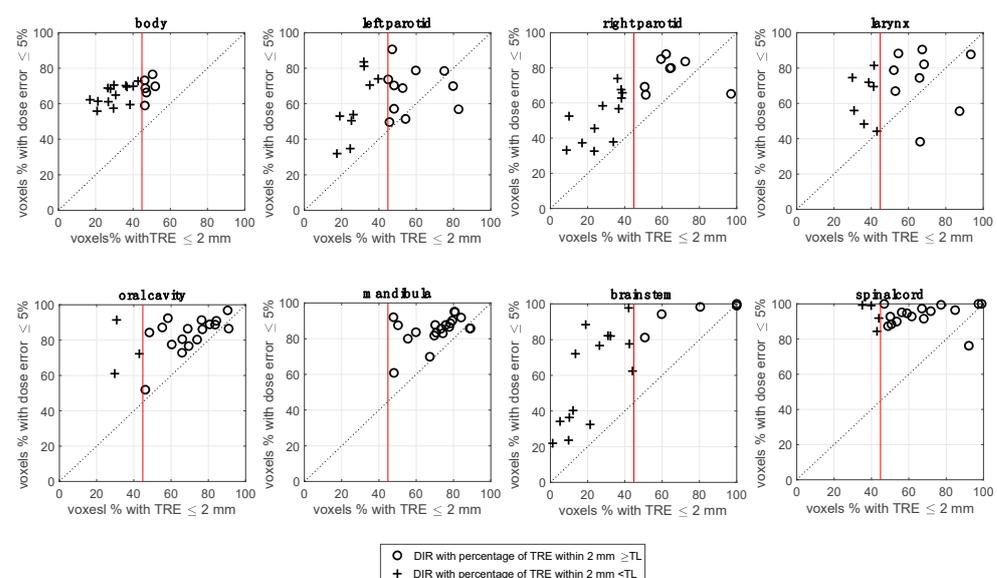


Figure 4. Correlation graph of the percentage of the TRE within 2 mm and percentage of voxels within 5% of dose error (3.5 Gy), across the 20 registrations and for each OAR. The '+' markers indicate the cases for which the percentage of voxels was less than the 45% OTL, indicated as vertical red line. The OTL value was derived using TG132 datasets, as described in Section 3.4.

3.5. Correlation Analysis

The correlation between the TRE and the magnitude of the deformation in the reference DVF was investigated on a voxel basis within the body contour, and a significant correlation (p -values < 0.01 across all patients) was found, with a Pearson correlation coefficient of 0.79 ± 0.14 (range 0.48–0.97), meaning that the DIR spatial errors are a function of the deformation magnitude of each registration.

As reported in the previously published studies [13,17,29], the effect of the DVF on the dose map is complex. Theoretically, it is expected to depend on the local dose gradient and on the spatial registration error [17]. The correlation between the dose map error and TRE was tested on an OAR-by-OAR basis, and the results are shown in Figure 5. The upper panel shows, in a correlation graph, the median value of the TRE versus the median value of the dose error for each OAR. The correlation values (indicated by the diameter of each marker size) varied significantly across the OARs.

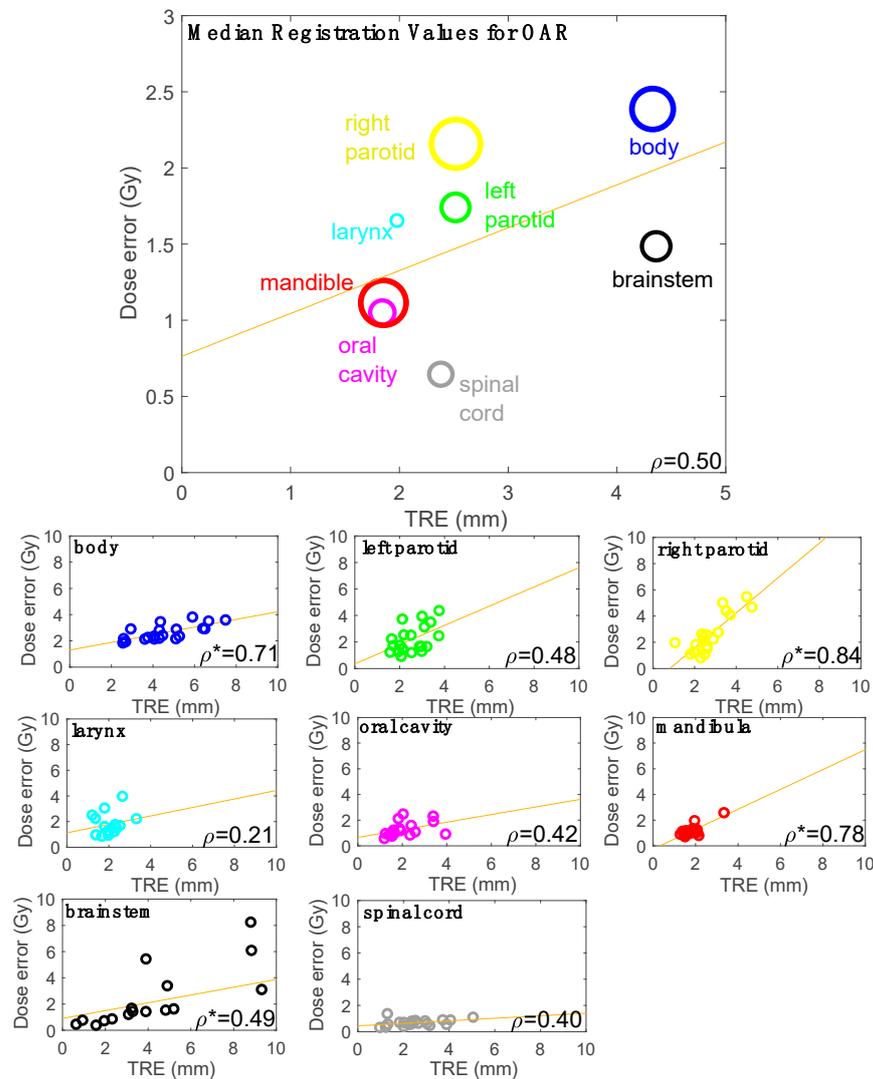


Figure 5. Upper panel: Correlation graph of the median TRE and dose errors values for each OAR. The diameter of each marker is proportional of the Pearson’s correlation coefficient value (see lower panels for individual numeric values). A linear correlation fit line is also reported. Lower panels: Correlation graph of the TRE and dose errors, across 20 registrations divided for each OAR. A linear correlation fit line is also reported. In the lower right corner of all panels, the Pearson’s coefficient was reported between TRE and dose error. A ‘*’ indicates that Pearson test confirmed statistically significant correlation.

The Figure 5 lower panels show, in correlation graphs, the TRE values versus the dose errors for each of the 20 registrations for each OAR. The Pearson coefficients, ranging from 0.21 (larynx) to 0.84 (right parotid), were reported for each panel. Significant correlations for the body, right parotid, mandibula and brainstem were found.

The correlation between the dose error and TRE was tested also with a voxel-based approach inside the body. A statistically significant correlation was found (p -values < 0.01), albeit with a low mean Pearson coefficient of 0.31 ± 0.10 (range 0.15–0.53) across all registrations [30]. This confirmed the complex correlation between the dose difference and TRE found on an OAR basis and reported above.

The dose error showed a moderate but statistically significant correlation with the dose gradient, with a mean Pearson coefficient of 0.51 ± 0.06 (range 0.37–0.61) and p -values < 0.01 across all patients on a voxel basis, confirming the hypothesis that the local dose gradient influences dose errors related to the DIR uncertainties, as reported also in the previous studies [17,31].

As found in the previous studies [13,17,29], the dose errors were also investigated as a function of the combination of the dose gradient and the TRE (here the scalar product between the two was considered). A highest mean Pearson coefficient of 0.52 ± 0.07 (range 0.41–0.66) and p -values < 0.001 were found, with respect to the single factors analysis (TRE and dose gradient).

4. Discussion

In this study, a novel method to evaluate the quality of patient-based multimodal CT-CBCT DIR using a post-processed, digitally enhanced anthropomorphic HN phantom has been presented. All post-processed ATOM Max Dental and Diagnostic Head Phantom image datasets (CT and CBCT) are in the DICOM format, derived using the procedure described in Section 2.1, and have been made publicly available as additional downloadable material for this paper as Supplementary Material. This allows the use of the proposed method for CT-CBCT HN DIR accuracy evaluation in other centres, by processing the image sets with two different image registration software systems (reference and clinical), as described in Section 2.1.

The refDVF represents the known transformation map derived with third-party reference software based on the clinically observed anatomical changes, which is then applied to create the warped phantom image. Thus, the transformation of each voxel in the warped image was known because it was deliberately applied to each voxel. As a verification method, the refDVF data did not necessarily represent the real, individual deformations of the patients, but they served as a reference to check how much the clinical algorithm under investigation (in this study, the Velocity AI) was able to reproduce the known deformation for that patient case. However, care must be taken to avoid applying a deformation derived using the same third-party algorithm that the user employs for the clinical deformation, as this would create a testDVF that was biased to the results expected from the same third-party and clinical DIR algorithms [6]. The registration algorithms used in this work to generate the refDVFs and testDVFs were both of the same class (B-spline) but implemented independently (i.e., different optimiser, similarity measure and general parameters). Because of these differences, the registrations can be considered independent between the two software programmes, in line with the previous studies [24,25].

The dose error variations were found to be related to both the registrations and the OAR of interest. This is in line with the findings of Qin et al. [29], who revealed that the impact of the DIR method on treatment dose warping is a function of the registration and is organ-specific. The OTL was derived from the material provided by the AAPM TG132 for quality assurance purposes. However, the TG132 utilised only pelvic anatomical data sets, processed using the ImSimQA, for the DIR accuracy assessment with a ground-truth transformation. Our findings confirmed that the registration accuracy is also a function of the selected organ at risk, underlying the importance of using an anatomical phantom for the DIR accuracy assessment, consistent with the clinical data considered. Thus, the OTL

derived for this study, using the generic pelvic phantom proposed by the TG132, could have some limitations on the practical level. However, it was used here primarily as an exemplar to describe a suggested workflow for the DIR accuracy, as described in Figure 1a.

In the published report on DIR validation by the AAPM's Task Group 132, the proposed ideal goal for the TRE is 95% of voxels within 2 mm, using the Report's provided image sets. Our results showed a considerably lower value, OTL = 45%, using the Velocity AI. The Velocity AI was previously benchmarked with different commercial DIR algorithms by Pukala et al. [13] and found to be very similar in accuracy with respect to the other commercial software systems included in the study. Based on this, it can be assumed that the OTL value found in this study could be considered a typical practical accuracy value using the commercial software, based on the B-spline method, for the CT-CBCT DIR registration using the TG-132 basic anatomical phantom. The difference between the OTL found in this study, and the goal suggested by the AAPM-TG132, underlines the pressing need for a universally available, comprehensive library of site-specific phantom datasets with the ground-truth deformation data to determine a more robust operative tolerance level to test the DIR software.

The registrations that pass the OTL for spatial accuracy can still result in a low percentage passing the dose difference criterion (as reported in Figures 4 and 5), indicating quite a complex relationship between the dose error and TRE accuracy. One limitation, shared by the spatial DIR metrics, is the indirect relationship between the quantification of the DIR error and the effect that error has on a given dose distribution. This is a complicated issue akin to invoking the gamma criterion for an IMRT QA analysis [32]. The TG-132 report did not suggest any tolerance goal for dose differences. For this reason, in the dosimetric analysis workflow reported in Figure 1, we have suggested a clinical evaluation criterion for the dose difference results.

A moderate, as classified in [30], but significant correlation on a voxel basis for both the dose gradient and the TRE with a dose difference was found. Veiga et al. [17] observed a moderate correlation between the dose gradient and the resulting dose difference, in agreement with our findings. Murphy et al. have reported the effect of the DVF on dose-mapping and concluded that there is a complex correlation that depends on the TRE and the voxel location, relative to the dose gradient [31]. The Pearson correlation test results were always significant ($p < 0.05$) for the voxel basis approach (Section 3.5). As reported by Schober et al. [33], the p -value derived from the test provides no information on how strongly the two variables are related. With large datasets, as they are in the voxel-based analysis, very small correlation coefficients can be "statistically significant." Therefore, a statistically significant correlation must not be confused with a clinically relevant correlation.

It is important to underline that this study represents a Proof of Principle, and further work will be undertaken to explore the behaviour of the algorithm in other situations. In our experience, the clinical workflow is composed of several tasks and can be time-consuming. Using the application conditions described (software and hardware), it can require up to 30 min for a registration evaluation. However, this time depends on several factors, such as the software involved, the type of registration DIR selected, the region of interest chosen, the VMAT plan parameters and the hardware characteristics for the registration software, the dose calculation and the experience of the operator. Moreover, a separate time allowance is required for the analysis of the DVFs and dose distribution at the voxel level, using the in-house software.

4.1. Range of Application of the Method

The use of a phantom in the method has some intrinsic limitations. The lack of an initial deformation or shift between the CT and CBCT (owing to the rigidity of the phantom and not applicable with the real patient CT and CBCT images) was the basis for the reference condition of the validity of the method. Since the clinical reference DVF has the same frame of reference as the patient CT image, care must be taken to align the DICOM coordinate

systems between the patient and HN phantom CT images before generating the warped phantom CT image based on the reference DVF. This was done to align the phantom CT image with the clinical refDVF to produce a more clinically realistic warped phantom CT image. Indeed, it is important that the warped phantom is created, applying as far as possible the deformation magnitude consistent with the spatial deformation of the patient images.

From a mathematical point of view, the validity of the method is consistent for every size of patient anatomy, because the quality of the registration was assessed as the ability of the clinical software to replicate the reference DVF, artificially introduced into the deformed phantom CT images. However, the rigid “coupling” between the phantom anatomy and the reference DVF should be considered carefully to ensure the clinical relevance of the reference DVF. The more the anatomy is similar between the phantom and patient, the more the clinical value of the reference DVF will be maintained when it is applied to the phantom anatomy. The quality of this coupling should be tested visually, by checking the overlap between the patient and phantom structures. A potential method to overcome the mismatch between the patient and phantom anatomy is to introduce a baseline DIR between the patient and the phantom in the coupling phase, instead of a rigid one. However, introducing this step can increase the complexity of the evaluation method, therefore it was decided to avoid this point to avoid over-complicating the workflow. Indeed, it is important to underline that, as a verification method, the refDVF data did not necessarily represent the real, individual deformations of the patients.

Moabbed et al. [34] used a physical anthropomorphic pelvic phantom for the CT and CBCT deformable image registration with the aim of validating a DIR evaluation method for prostate cancer cases. Even if their study focused on evaluating the accuracy of the dose calculation as performed directly on the CBCT images, they presented a similar workflow to the current study for their evaluation, in that the pelvic phantom images were digitally deformed using two realistic patient-based deformation fields, after a rigid alignment between the patient and the phantom’s system of reference.

Since the OAR insertion in the HN processed phantom was done in terms of HU, and not in physical density, the processed phantom images were not suited to studies of the dose calculation accuracy, but only for the DIR registration accuracy. Indeed, the dose calculation using this method is only performed on the patient adapted-CT.

4.2. Similar DIR Accuracy Workflows to the Current Study

Our proposed method quantified the effect of the DIR on the dose distribution, focusing on the DVF accuracy. As was reported in the introduction, for adaptive radiation therapy, the accurate dose deformation and summation/accumulation within each contour is determined by the DVF accuracy [16]. Other methods evaluated the dose deformation accuracy, focusing on the DVF.

Recently, Lowther et al. [35] assessed the uncertainty in the dose accumulation procedure for the HN adaptive by comparing the DIR-facilitated dose accumulation using a commercial software system with the results of an in-silico model, based on the clinically observed deformations as the ground truth. HN patients CT and CBCTs images were used to generate in silico reference CBCTs and DVFs, as the ground truth, with a second reference B-spline DIR software. The in silico images, generated with an independent DIR software as the reference, were necessary to have a known reference DVF. Also, this method has the advantage of evaluating each patient-specific deformation and dose warping. However, Lowther’s method works directly on the patients’ images, making the clinical workflow simpler and reasonably quicker than the proposed one, avoiding the use of the HN phantom presented in this study, with its related inherent imitations (see Section 4.1). The complexity of the deformations in the human body, as well as its heterogeneity in terms of the HU values, are challenging to model accurately through a phantom. On the other hand, the method implemented in our work has the advantage of replicating the HU distribution found in the OARs on clinical images. In comparison, Lowther’s method tested the DIR

between the planning CT and the in silico model of the CBCT, which is different from the original CBCT in the HU and quality (the planning CT deformed to the anatomy of the CBCT of the considered fraction) and could not consider some issues of the CTBT-to-CT DIR, including the reduction in the accuracy due to the low image contrast of the real CBCT.

Pukala et al. [13] evaluated the performance of several DIR software systems using several HN digital phantoms, in terms of both the spatial and dosimetric accuracy at the voxel level. The study created publicly available phantoms based on the CT data from head and neck cancer patients. These phantoms provide a clinically based ground-truth model that encompasses the anatomical changes that occur over the course of a typical treatment. Also, this study assessed the deformation accuracy for each voxel by comparing the entire registration DVF to the ground-truth DVF, based on the clinical scenario. One key aspect of our study was the possibility to use real patient data (the DVF and dose distributions) to evaluate each patient-specific DIR, instead of commissioning specific DIR algorithms. Moreover, the study of Pukala was focused on CT-to-CT registrations, while our study referred to multimodal CBCT-to-CT registration. The latter presents more issues in the DIR accuracy, due to the limited quality of the CBCT images.

Veiga et al. [17] investigated the transformations that mapped the anatomy between the planning CT and CBCT using four different DIR approaches. DVFs were used to remap the dose of the day onto the planning CT. The data from five HN patients were used to evaluate the performance of each implementation, based on geometrical matching, the physical properties of the DVFs, and the similarity between the warped dose distributions. As reported in the Introduction, this approach can be useful to choose the optimal registration method, without conveying any information on the DIR uncertainties and related dose errors.

Singhrao et al. [36] used a deformable HN phantom to test DIR algorithms, but, in practice, it is difficult to produce physical phantoms for every deformation scenario that occurs clinically.

Essentially, the novelty of our approach consists of the use of an HN anthropomorphic phantom for the CT-CBCT registration, which was digitally enhanced to include the OARs which are typical of that anatomic region, (Figure 1), with a robust accuracy verification procedure taking into account the recommended OARs for the HN. Also, in this study, further recommendations in terms of a clinically feasible workflow and achievable thresholds for the CBCT-to-CT deformable image registration have been investigated. The findings of this study indicate that there is a need for standardised approaches and specific guidelines for different applications to supplement the overarching DIR guidelines such as the TG132.

5. Conclusions

A novel method to evaluate the quality of patient-specific multimodality CT-CBCT DIR for the adaptive radiotherapy of HN patients, based on the clinical observed changes as the ground truth, has been developed, using a digitally enhanced anthropomorphic phantom to include the common organs at risk. An established workflow and operative tolerance level, based on the TG132 indications, were provided. The methodology allows the direct testing of the patient-specific DIR spatial and dosimetric accuracy at the voxel level, which can produce valuable insights into their clinical impact on the adapted dose distribution.

The method was applied to test the DIR performance on a set of clinical HN patients using clinical commercial software. It was found that the DIR performance varied with the specific registration and organ, and the complex relationship between the spatial and dose errors. The results of this study suggest that the method is suitable to assess the adequacy of the DIR-based dose accumulation.

The technical requirement for using this method is the availability of an independent image registration software platform (in addition to the clinical system) and the processed phantom image sets. This work contributes to the study of the standardisation and automation of quality assurance methods for deformable image registration in radiotherapy.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12199493/s1>.

Author Contributions: Conceptualization, E.C., A.B. and M.I.; methodology, E.C., A.B., G.L. and E.S.; investigation, E.C. and A.B.; resources, M.O. and M.G.; data curation, E.C., A.B., G.L. and E.S.; writing—original draft preparation, E.C., G.L. and E.S.; writing—review and editing, G.L. and E.S.; supervision, C.I., M.I., E.S. and G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by Italian Ministry of Health–Ricerca Corrente.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Azienda USL-IRCCS di Reggio Emilia (with protocol code n. 2016/0014486, Date 30 May 2016).

Informed Consent Statement: Patient consent was waived because of the retrospective design of the study and to the impossibility of providing the informed consent statement to each subject involved in the protocol study, due to the high number of interested parties (Italian law, Legislative Decree 196/2003, Art.110 and Deliberazione del Garante Privacy n. 85 on 1 March 2012 (published in the GU on 26 March 2012)).

Data Availability Statement: The data are available from the corresponding author.

Acknowledgments: The authors would like to thank ELSE Solutions (Trezzano (MI), Italy) and CIRS (Norfolk, VA, USA) for providing the ATOM Max TM Dental and Diagnostic Head Phantom Model-711 used for this study. The authors would also like to thank Giacomo Bartesaghi for his technical support and Agnese Chendi for her useful comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Keall, P.; Poulsen, P.; Booth, T. See, Think and Act: Real-Time Adaptive Radiotherapy. *Semin. Radiat. Oncol.* **2019**, *29*, 228–235. [[CrossRef](#)] [[PubMed](#)]
2. Jaffray, D.A. Image-guided radiotherapy: From current concept to future perspectives. *Nat. Rev. Clin. Oncol.* **2012**, *9*, 688–699. [[CrossRef](#)] [[PubMed](#)]
3. Liu, C.; Kim, J.; Kumarasiri, A.; Mayyas, E.; Brown, S.L.; Wen, N.; Siddiqui, F.; Chetty, I.J. An automated dose tracking system for adaptive radiation therapy. *Comput. Methods Programs Biomed.* **2018**, *154*, 1–8. [[CrossRef](#)] [[PubMed](#)]
4. Heukelom, J.; Fuller, C.D. Head and Neck Cancer Adaptive Radiation Therapy (ART): Conceptual Considerations for the Informed Clinician. *Semin. Radiat. Oncol.* **2019**, *29*, 258–273. [[CrossRef](#)] [[PubMed](#)]
5. Green, O.L.; Henke, L.E.; Hugo, G.D. Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy. *Semin. Radiat. Oncol.* **2019**, *29*, 219–227. [[CrossRef](#)] [[PubMed](#)]
6. Brock, K.K. *Image Processing in Radiation Therapy*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2013. [[CrossRef](#)]
7. Kadoya, N.; Nakajima, Y.; Saito, M.; Miyabe, Y.; Kurooka, M.; Kito, S.; Fujita, Y.; Sasaki, M.; Arai, K.; Tani, K.; et al. Multi-institutional validation study of commercially available deformable image registration software for thoracic images. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *96*, 422–431. [[CrossRef](#)] [[PubMed](#)]
8. Loi, G.; Fusella, M.; Lanzi, E.; Cagni, E.; Garibaldi, C.; Iacoviello, G.; Lucio, F.; Menghi, E.; Miceli, R.; Orlandini, L.C.; et al. Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: A multi-institutional study. *Med. Phys.* **2018**, *45*, 748–757. [[CrossRef](#)] [[PubMed](#)]
9. Viergever, M.A.; Maintz, J.A.; Klein, S.; Murphy, K.; Staring, M.; Pluim, J.P. A survey of medical image registration—under review. *Med. Image Anal.* **2016**, *33*, 140–144. [[CrossRef](#)] [[PubMed](#)]
10. Brock, K.K.; Mutic, S.; McNutt, T.R.; Li, H.; Kessler, M.L. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med. Phys.* **2017**, *44*, e43–e76. [[CrossRef](#)] [[PubMed](#)]
11. Rong, Y.; Rosu-Bubulac, M.; Benedict, S.H.; Cui, Y.; Ruo, R.; Connell, T.; Kashani, R.; Latifi, K.; Chen, Q.; Geng, H.; et al. Rigid and Deformable Image Registration for Radiation Therapy: A Self-Study Evaluation Guide for NRG Oncology Clinical Trial Participation. *Pract. Radiat. Oncol.* **2021**, *11*, 282–298. [[CrossRef](#)] [[PubMed](#)]
12. Latifi, K.; Caudell, J.; Zhang, G.; Hunt, D.; Moros, E.G.; Feygelman, V. Practical quantification of image registration accuracy following the AAPM TG-132 report framework. *J. Appl. Clin. Med. Phys.* **2018**, *19*, 125–133. [[CrossRef](#)] [[PubMed](#)]
13. Pukala, J.; Johnson, P.; Shah, A.P.; Langen, K.M.; Bova, F.J.; Staton, R.J.; Mañon, R.R.; Kelly, P.; Meeks, S.L. Benchmarking of five commercial deformable image registration algorithms for head and neck patients. *J. Appl. Clin. Med. Phys.* **2016**, *17*, 25–40. [[CrossRef](#)]

14. Hussein, M.; Akintonde, A.; McClelland, J.; Speight, R.; Clark, C.H. Clinical use, challenges, and barriers to implementation of deformable image registration in radiotherapy—the need for guidance and QA tools. *Br. J. Radiol.* **2021**, *94*, 20210001. [[CrossRef](#)] [[PubMed](#)]
15. Cinaglia, P.; Tradigo, G.; Cascini, G.L.; Zumpano, E.; Veltri, P. A framework for the decomposition and features extraction from lung DICOM images. In Proceedings of the 22nd International Database Engineering & Applications Symposium, Villa San Giovanni, Italy, 18–20 June 2018; pp. 31–36. [[CrossRef](#)]
16. Shi, L.; Chen, Q.; Barley, S.; Cui, Y.; Shang, L.; Qiu, J.; Rong, Y. Benchmarking of Deformable Image Registration for Multiple An-atomic Sites Using Digital Data Sets With Ground-Truth Deformation Vector Fields. *Pract. Radiat. Oncol.* **2021**, *11*, 404–414. [[CrossRef](#)] [[PubMed](#)]
17. Veiga, C.; Lourenço, A.M.; Mouinuddin, S.; Van Herk, M.; Modat, M.; Ourselin, S.; Royle, G.; McClelland, J.R. Toward adaptive radiotherapy for head and neck patients: Uncertainties in dose warping due to the choice of deformable registration algorithm. *Med. Phys.* **2015**, *42*, 760–769. [[CrossRef](#)]
18. Veiga, C.; McClelland, J.; Moinuddin, S.; Lourenço, A.; Ricketts, K.; Annkah, J.; Modat, M.; Ourselin, S.; D'Souza, D.; Royle, G. To-ward 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. [[CrossRef](#)] [[PubMed](#)]
19. Paganelli, C.; Meschini, G.; Molinelli, S.; Riboldi, M.; Baroni, G. Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats. *Med. Phys.* **2018**, *45*, e908–e922. [[CrossRef](#)] [[PubMed](#)]
20. Hatton, J.; McCurdy, B.; Greer, P.B. Cone beam computerized tomography: The effect of calibration of the Hounsfield unit number to electron density on dose calculation accuracy for adaptive radiation therapy. *Phys. Med. Biol.* **2009**, *54*, N329–N346. [[CrossRef](#)] [[PubMed](#)]
21. Schulze, R.; Heil, U.; Grob, D.; Bruellmann, D.D.; Dranischnikow, E.; Schwanecke, U.; Schoemer, E. Artefacts in CBCT: A review. *Dentomaxillofac. Radiol.* **2011**, *40*, 265–273. [[CrossRef](#)]
22. Brouwer, C.L.; Steenbakkers, R.J.; Bourhis, J.; Budach, W.; Grau, C.; Grégoire, V.; van Herk, M.; Lee, A.; Maingon, P.; Nutting, C.; et al. CT-based delineation of organs at risk in the head and neck region: DA-HANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol.* **2015**, *117*, 83–90. [[CrossRef](#)]
23. TrueBeam Technical Reference Guide—Volume 2: Imaging 2021. Available online: www.myvarian.com (accessed on 20 March 2020).
24. Deasy, J.O.; Blanco, A.I.; Clark, V.H. CERR: A computational environment for radiotherapy research. *Med. Phys.* **2003**, *30*, 979–985. [[CrossRef](#)] [[PubMed](#)]
25. Brouwer, C.L.; Steenbakkers, R.J.; Langendijk, J.A.; Sijtsema, N.M. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy pro-vide information to help? *Radiat. Oncol.* **2015**, *115*, 285–294. [[CrossRef](#)] [[PubMed](#)]
26. Klein, S.; Staring, M.; Murphy, K.; Viergever, M.A.; Pluim, J.P. Elastix: A toolbox for intensity-based medical image registration. *Trans. Med. Imaging* **2010**, *29*, 196–205. [[CrossRef](#)]
27. Broggi, S.; Scalco, E.; Belli, M.L.; Logghe, G.; Verellen, D.; Moriconi, S.; Chiara, A.; Palmisano, A.; Mellone, R.; Fiorino, C.; et al. A Comparative Evaluation of 3 Different Free-Form Deformable Image Registration and Contour Propagation Methods for Head and Neck MRI: The Case of Parotid Changes During Radiotherapy. *Technol. Cancer Res. Treat.* **2017**, *16*, 373–381. [[CrossRef](#)]
28. Velocity AI v.4.1 Instruction for Use, Version 4.1; Varian Medical System. Available online: www.myvarian.com (accessed on 20 September 2020).
29. Qin, A.; Liang, J.; Han, X.; O'Connell, N.; Yan, D. Technical Note: The impact of deformable image registration methods on dose warping. *Med. Phys.* **2018**, *45*, 1287–1294. [[CrossRef](#)] [[PubMed](#)]
30. Asuero, A.G.; Sayago, A.; González, A.G. The Correlation Coefficient: An Overview. *Crit. Rev. Anal. Chem.* **2006**, *36*, 41–59. [[CrossRef](#)]
31. Murphy, M.J.; Salguero, F.J.; Siebers, J.V.; Staub, D.; Vaman, C. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Med. Phys.* **2012**, *39*, 573–580. [[CrossRef](#)] [[PubMed](#)]
32. Low, D.A.; Harms, W.B.; Mutic, S.; Purdy, J.A. A technique for the quantitative evaluation of dose distributions. *Med. Phys.* **1998**, *25*, 656–661. [[CrossRef](#)]
33. Schober, P.; Boer, C.; Schwarte, L. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)]
34. Moteabbed, M.; Sharp, G.C.; Wang, Y.; Trofimov, A.; Efsthathiou, J.A.; Lu, H.M. Validation of a deformable image registration technique for cone beam CT-based dose verification. *Med. Phys.* **2015**, *42*, 196–205. [[CrossRef](#)]
35. Lowther, N.J.; Marsh, S.H.; Louwe, R.J. Quantifying the dose accumulation uncertainty after deformable image registration in head-and-neck radiotherapy. *Radiother. Oncol.* **2020**, *143*, 117–125. [[CrossRef](#)] [[PubMed](#)]
36. Singhrao, K.; Kirby, N.; Pouliot, J. A three-dimensional head-and-neck phantom for validation of multimodality deformable image registration for adaptive radiotherapy. *Med. Phys.* **2014**, *41*, 121709. [[CrossRef](#)]