However, the assessment of pseudoCT-derived errors should also be carried out in other treatment sites where dose to critical structures will guide dose escalation and adaptive RT strategies, such as isotoxic escalation in lung and head and neck RT.

PO-1752 Evaluation of the deformable image registration algorithm in Velocity image registration software

R. Fullarton[1], K. Ghirmay[1], W. Dom[1], L. Crees[1]
[1]*Queen's Hospital, Radiotherapy Physics, Romford, United Kingdom*

Purpose or Objective

Deformable image registration (DIR) algorithms can be used to support clinical decisions for adaptive radiotherapy strategies. The spatial accuracy of the algorithms can indicate the feasibility of clinical application of the DIR, in the field of adaptive radiotherapy, for planning and delineating the target and organ-at-risk volumes accurately. The aim of this work is to evaluate the accuracy of the DIR algorithm within the Velocity image registration software (Varian, Pasadena, CA) for different treatment sites.

Material and Methods

The planning CT of ten patients, who underwent radiotherapy treatments for head & neck, thorax and pelvic-based cancers, were retrospectively selected. Deformations, for which 3D transformation is known via the Deformation Vector Field (DVF), were applied to the patient CT datasets. The deformed CT dataset was then transferred into the Velocity software for re-registration with the unmodified patient CT dataset. The DVF provided by the Velocity software was compared to the known DVF, to quantify the accuracy of the deformable algorithm, on a voxel by voxel basis. The report from AAPM TG-132 [1] recommends that the mean vector difference from the known transformation should be less than 2.00 mm.

Results

The mean and median vector differences of the vector fields for the thorax and head & neck dataset registrations were less than 2.00 mm for all datasets. However, this was not achieved for the pelvis patient datasets for which three registrations were found to have a mean vector difference greater than 2.00 mm (2.19-2.43mm). The mean vector difference for each site was found to be 1.53mm, 1.28mm and 1.08mm for the pelvis, head & neck and thorax cases respectively. A Kruskal-Wallis one-way analysis of variance (ANOVA), at significance level $p<0.05$, was carried out to verify the differences in performance between the scanned sites. The results indicate that the differences in mean and median vector differences between the sites were found to be significant. A further investigation was also carried out to check that the variances were not due to a bias in the initial deformations. ANOVA tests were run on the mean and median values of the initial vector field and the tests showed no significant difference ($p<0.05$).

Conclusion

The DIR algorithm within Velocity provided good agreement with the known deformations across all sites with the best agreement being observed in the thorax cases and the worst agreement being observed in the pelvis cases. This difference in registration accuracy could be due to the more homogeneous HU values in the pelvis compared to the lung and head & neck sites leading to the two algorithms deforming the anatomy differently. The difference between the sites may help inform uncertainties and planning margins when using the software for adaptive radiotherapy.

1. Brock K, Mutic S, McNutt T, Li H, Kessler M. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Task Group No. 132. 2019.

PO-1753 Full-body delineation of ROIs through anatomy-preserving deep learning ensemble networks

A. Lombard[1], L. De Vitry D'Avaucourt[1], E. Ullmann[1], R. Marini-Silva[1], N. Bus[1], N. Paragios
[1]*Therapanacea, Artificial Intelligence, Paris, France*

Purpose or Objective

Delineation of anatomical structures is a tedious/time-consuming task of radiation therapy associated with strong inter/intra-user variability. In this study, we introduce an automatic tool that detects and segments 65+ anatomical structures including lymph nodes. Our approach simultaneously (i) detects the organs present in the volume, (ii) delineates them through a combination of ensemble of neural networks, (iii) introduces anatomical consistency though a segmentation by deformation principle.

Material and Methods

The data set contains 22,000 volumes (after data augmentation) from several clinical sites. The annotations per organ varied; 680 for the cervical lymph nodes level VIIA to 12,232 for the spinal cord. We deploy a full end-to-end ensemble neural networks approach involving three steps: (i) detection of the organs present in the volume through a deep neural network that registers the volume to an ensemble of whole-body annotated volumes, (ii) automatic delineation of each anatomical structure through a unique combination of data-driven & decisional artificial intelligence & (iii) a winner takes all approach and enforces anatomical consistency. Multiple networks are trained using different whole body scans as referenced space. Each of them relies on a different random separation between training (80%) and validation (20%) subsets. Evaluation was done using a random representative set of the data set, while the method was also tested on multicentric external cohorts - results presented in different submissions - (Head and Neck & Breast anatomies).

Results

Mean Dice scores (MDS), mean contour distance (MCD) and mean Hausdorff distance (MHD) were evaluated for all anatomical structures. Encephalon was the structure with the best performance (MDS: 0.99, MCD/MHD: 0.23mm/5.5mm) while chiasma the worst (MDS: 0.45, MCD/MHD: 1.24/6.83), a structure that inherits strong difficulties even for clinical experts when contoured on a CT. Running time is approximately 3 seconds per anatomical structure present in the image. The average dice score for all organs was 0.79, the standard deviation 0.10. More than 50% of the structures had a MDS above 0.8 and only 4 among them under 0.6. Representative results on a Head and Neck and Prostate cases as well as dice scores for a randomly selected subset of organs are presented:

Organ	MDS (range)
Eyes	0.91 ±0.04
Optical nerves	0.66 ±0.06
Parotids	0.803 ±0.12
Submandibles	0.83 ±0.17
Cervical lymph nodes level III	0.79 ±0.05
Esophagus	0.72 ±0.15
Breast	0.88 ±0.15
Heart	0.92 ±0.03
Bladder	0.93 ±0.07
Prostate	0.88 +-0.04
Seminal vesicles	0.74 +-0.15
CTVN Prostate	0.82 +-0.07

Conclusion

We introduced a fully automatic deep learning ensemble network approach that couples organ detection & anatomy-preserving annotation. The obtained results are highly promising – exceeding human precision in a number of cases – and consist of a prominent avenue for automatization, standardization and healthcare cost reduction in radiation therapy.

PO-1754 A stereoscopic CT artifact reduction method image quality comparison to current vendor solutions

D. Branco[1], S. Kry[1], T. Paige[1], J. Rong[2], X. Zhang[1], S. Frank[3], D. Followill[1]
[1]MD Anderson Cancer Center, Radiation Physics, HOUSTON, USA ; [2]MD Anderson Cancer Center, Imaging Physics, Houston, USA ; [3]MD Anderson Cancer Center, Radiation Oncology, Houston, USA

Purpose or Objective

The purpose of this work was to perform a quantitative image quality comparison of a novel stereoscopic in-house developed CT head and neck metal artifact reduction (MAR) technique (AMPP) to current vendors' commercial algorithms.

Material and Methods

An anthropomorphic phantom composed of tissue equivalent materials, a human skull and air cavities was used. The jaw insert was created to allow for obtaining images with and without (baseline) metal fillings. The phantom was scanned using Philips, Siemens, GE and Toshiba CT scanners where each metal scan was reconstructed with its respective vendor's MAR algorithm so that each proprietary solution was evaluated. All algorithms were evaluated for severity of artifacts and CT number accuracy. HU number accuracy was quantified for each vendor's corrected scan and its respective baseline scan where the mean HU number differences and standard deviations inside contoured volumes were obtained. To quantify the severity of artifacts, HU difference maps and percentage of bad pixels were obtained where all pixels with an HU error outside ±20 HU were considered to be bad pixels. In addition, the in-house algorithm was also evaluated for robustness using different imaging parameters.

Results

Each of the vendor's algorithms generally improved the severity of artifacts found in the uncorrected image set. It is also noticeable that the streaking is still visible in the commercial solutions and, in some instances, introduces additional artifacts in the posterior region, in contrast to AMPP which is clear of artifacts and nearly identical to the baseline image posterior to the oral cavity (Fig 1). For the volumetric HU analysis, AMPP consistently outperformed the other MAR algorithms and improved the HU accuracy to nearly the same as the uncorrected baseline scan. The differences shown in the vendors MARs algorithms difference maps are improved, but still show remarkable inconsistencies with their baselines. In comparison, AMPP shows great agreement with its baseline (within ±20 HU) displaying a mostly green difference map inside and outside the phantom (Fig 2). The percent of bad pixels in the circular region of the cylinder was 78.1%, 65.5%, 29.1%, 25.5%, 27.9%, 4.2% for the uncorrected, OMAR, SEMAR, iMAR, SmartMAR and AMPP scans, respectively. For the robustness analysis, AMPP performed similarly across the different scan techniques despite the varying imaging parameters chosen; different AMPP scans showed similarly small HU differences and percent of bad pixels compared to each respective baseline.