

Segmentation-Guided Multi-Modal Registration of Liver Images for Dose Estimation in SIRT

Xikai Tang^{1,7}, Esmaeel Jafargholi Rangraz⁹, Richard ‘s Heeren¹, Walter Coudyzer⁵, Geert Maleux^{3,5}, Kristof Baete^{1,4}, Chris Verslype⁶, Mark J. Gooding⁸, Christophe M. Deroose^{1,4}, Johan Nuyts^{1,7}

¹Nuclear Medicine and Molecular Imaging, KU Leuven, Leuven, Belgium; ²Medical Image Computing (ESAT/PSI), KU Leuven, Leuven, Belgium; ³Radiology, KU Leuven, Leuven, Belgium; ⁴Nuclear Medicine, University Hospitals Leuven, Leuven, Belgium; ⁵Radiology, University Hospitals Leuven, Leuven, Belgium; ⁶Digestive Oncology, University Hospitals Leuven, Leuven, Belgium; ⁷Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium. ⁸Mirada Medical Ltd, Oxford, UK; ⁹Quirem Medical B.V., Deventer, Netherlands.

For correspondence contact and the first author (currently in training, doctoral student):

Name: Xikai Tang

Address: Medical Imaging Research Center, UZ Herestraat 49 - box 7003, 3000 Leuven, Belgium.

Telephone number: +32 16 3 49084

Fax number: +32 16 3 43759

Email: xikai.tang@kuleuven.be

ORCID: 0000-0002-9877-0294

Acknowledgement: This project is funded by the H2020-ITN (MSCA 764458) project Hybrid and by the Research Foundation Flanders (FWO) project G082418N. We thank Jeroen Bertels and David Robben for providing the code of the CNN structure. We thank Georg Schramm for providing the tool for image format conversion. We thank Wies Deckers for providing the technical support.

EJNMMI Phys **9**, 3 (2022). <https://doi.org/10.1186/s40658-022-00432-8>

Abstract

Purpose: Selective internal radiation therapy (SIRT) requires a good liver registration of multi-modality images to obtain precise dose prediction and measurement. This study investigated the feasibility of liver registration of CT and MR images, guided by segmentation of the liver and its landmarks. The influence of the resulting lesion registration on dose estimation was evaluated. **Methods:** The liver segmentation was done with a convolutional neural network (CNN) and the landmarks were segmented manually. Our image-based registration software and its liver-segmentation-guided extension (CNN-guided) were tuned and evaluated with 49 CT and 26 MR images from 20 SIRT patients. Each liver registration was evaluated by the root mean square distance (RMSD) of mean surface distance between manually delineated liver contours and mass center distance between manually delineated landmarks (lesions, clips, etc). The root mean square of RMSDs (RRMSD) was used to evaluate all liver registrations. The CNN-guided registration was further extended by incorporating landmark segmentations (CNN&LM-guided) to assess the value of additional landmark guidance. To evaluate the influence of segmentation-guided registration on dose estimation, mean dose and volume percentages receiving at least 70 Gy (V70) estimated on the ^{99m}Tc-labeled macro-aggregated albumin (^{99m}Tc-MAA) SPECT were computed, either based on lesions from the reference ^{99m}Tc-MAA CT (reference lesions) or from the registered floating CT or MR images (registered lesions) using the CNN- or CNN&LM-guided algorithms. **Results:** The RRMSD decreased for the floating CTs and MRs by 1.0 mm (11%) and 3.4 mm (34%) using CNN guidance for the image-based registration and by 2.1 mm (26%) and 1.4 mm (21%) using landmark guidance for the CNN-guided registration. The quartiles for the relative mean dose difference (the V70 difference) between the reference and registered lesions and their correlations [25th, 75th; r] are as follows: [-5.5% (-1.3%), 5.6% (3.4%); 0.97 (0.95)] and [-12.3% (-2.1%), 14.8% (2.9%); 0.96 (0.97)] for the CNN&LM- and CNN-guided CT to CT registrations, [-7.7% (-6.6%), 7.0% (3.1%); 0.97 (0.90)] and [-15.1% (-11.3%), 2.4% (2.5%); 0.91 (0.78)] for the CNN&LM- and CNN-guided MR to CT registrations. **Conclusion:** Guidance by CNN liver segmentations and landmarks markedly improves the performance of the image-based registration. The small mean dose change between the reference and registered lesions demonstrates the feasibility of applying the CNN&LM- or CNN-guided registration to volume-level dose prediction. The CNN&LM- and CNN-guided registrations for CTs can be applied to voxel-level dose prediction according to their small V70 change for most lesions. The CNN-guided MR to CT registration still needs to incorporate landmark guidance for smaller change of voxel-level dose estimation.

Key Words: selective internal radiation therapy (SIRT); liver registration; convolutional neural network (CNN); internal dosimetry; multi-modality images

I. Introduction

Selective internal radiation therapy (SIRT) or radioembolization is increasingly applied for the treatment of surgically unresectable primary liver malignancies and secondary metastases. During this treatment, microspheres loaded with β -emitting radionuclides, including yttrium-90 (^{90}Y) or holmium-166 (^{166}Ho), are infused into selected branches of the hepatic artery according to the vascular anatomy mapped by angiography [1]. Since the selected branches dominate the blood supply to tumors [2], these radioactive microspheres are trapped within the tumors. The high energy, small tissue penetrating range, and concentration in tumors allow these microspheres to deposit higher energy per mass in tumors, while preventing healthy liver parenchyma dysfunction by limiting their irradiation..

In the SIRT planning, the absorbed dose is used to measure the amount of energy per mass (in Gy or J/kg) from ionizing radiation deposited in a volume of interest (VOI), including tumors and healthy liver parenchyma. It serves as a toxicity indicator for tumors and normal tissues and a criterion for determining the amount of injected activity. There are several different methods to determine the injected activity, including mono-compartment [3] and multi-compartment [4] methods and voxel-based approaches [3]. The dose calculation better reflects the underlying biology as the VOI changes from the volume level (the whole liver, tumors and non-tumoral parts) to the voxel level.

During the pre- and post-treatment studies, multi-modality images are acquired for VOI delineation and dose calculation. In the pre-treatment study, a SPECT/CT scan is performed shortly after administration of $^{99\text{m}}\text{Tc}$ -labeled macro-aggregated albumin ($^{99\text{m}}\text{Tc}$ -MAA) particles into selected branches of hepatic artery to mimic the activity distribution of ^{90}Y microspheres in the liver [5]. Cone beam CT (CBCT) images are used for delineation of liver perfusion territories (LPTs) [6]. ^{18}F -fluorodeoxyglucose ($[^{18}\text{F}]\text{FDG}$) or $[^{68}\text{Ga}]\text{DOTATATE}$ PET/CT scans are performed for $[^{18}\text{F}]\text{FDG}$ - or $[^{68}\text{Ga}]\text{DOTATATE}$ -avid tumors. Contrast-enhanced and diffusion-weighted MR image are acquired for tumors that are not $[^{18}\text{F}]\text{FDG}$ - or $[^{68}\text{Ga}]\text{DOTATATE}$ -avid. Following the administration of ^{90}Y -microspheres, a post-treatment study is performed to obtain the PET/CT or PET/MR images of the actual activity distribution inside the liver [7]. This PET image can be used to compute the absorbed dose in different VOIs for evaluation of the treatment irradiation distribution.

Registration of multi-modality images plays an essential role in information integration for SIRT dosimetry. Different methods of liver registration for intra-modality (CT or MR) and inter-modality images (CT and MR) have been studied. These methods include surface-based [8-11], vessel-based [9, 11, 12], intensity-based [10, 13, 14], and segmentation-based registrations [15, 16]. Some studies combine both liver surfaces and vascular structures for better registration of tumors inside the liver [9, 11], since most tumors are found near vessels. Most intensity-based methods adopt mutual information for image similarity measurement. The segmentation-based methods use liver segmentations obtained either with histogram-based thresholding [15] or with a convolutional neural network (CNN) [16] for guidance of rigid or affine registration. For the application of liver registration in SIRT, Alsultan *et al.* used a rigid registration method in Simplicit90Y (Mirada Medical Ltd, Oxford, UK) to register contrast-enhanced CT images to low-dose CT images for efficacy evaluation of coil embolization to acquire intrahepatic redistribution [17]. Spahr *et al.* implemented a registration framework based on normalized gradient fields for liver registration of multi-modality images, and evaluated their algorithm through landmarks and deformation field analyses [18]. Nodari *et al.* used a multi-modality deformable registration algorithm in MIM SurePlan (v7.0.1; MIM software Cleveland, USA), performed by a trained medical physicist, to register the liver tumor contours from MR images to $^{99\text{m}}\text{Tc}$ -MAA-SPECT/CT and ^{90}Y -PET/CT [19]. Their study indicates that different tumor contours from anatomical and scintigraphic images have no significant impact on mean dose, and registering anatomical tumor contours to scintigraphic images is feasible for improving therapeutic strategy [19]. Besides, our in-house non-rigid liver registration regularized by a spring model was applied in our previous studies on the development of pre-treatment dosimetry [6] and evaluation of the predictive value of $^{99\text{m}}\text{Tc}$ -MAA-based dose planning [20].

This study describes a (semi-)automatic segmentation-guided registration algorithm, and evaluates its performance for registering liver contours and anatomic landmarks and its influence on dose estimation for SIRT. Our in-house image-based registration algorithm was modified for guidance by CNN liver segmentations. After that, manually delineated landmarks (lesions, vessels, clips, etc.), together with CNN liver segmentations, were used for registration guidance to explore the possibility of using automatic lesion segmentations to improve lesion registration in the future. In this study, all CT and MR images from the pre- and post-treatment studies are registered to $^{99\text{m}}\text{Tc}$ -MAA CTs. Lesion contours were manually delineated on each CT and MR image for each patient independently. Dose estimations, based on lesion contours defined either on the $^{99\text{m}}\text{Tc}$ -MAA CT or on registered images, are compared to evaluate the influence of the registration algorithm on dosimetry. (Semi-)automatic registration algorithms have the advantage of providing the physicians with integrated image information for more precise dosimetry, while not creating much cumbersome and time-consuming work in the clinical workflow. In our experience, the current lack of automation impedes the full use of multi-modality information. The registration can be challenging, because in the $^{99\text{m}}\text{Tc}$ -MAA SPECT/CT protocol, the CT is acquired without contrast enhancement, and some MR images from the ^{90}Y -PET/MR studies have severe artefacts. Through this study, we want to implement (semi-)automatic liver registration to demonstrate its feasibility and value in clinical context, and contribute to a personalized and precise SIRT treatment with fewer manual interactions.

II. Material and methods

A. Data

The training datasets for the CNN model contain 119 CT images from Liver Tumor Segmentation (LiTS17) and SLIVER07 challenges, 30 MR T1 images from the CHAOS challenge, and 50 SIRT CT and 110 MR T1 images from our hospital.

For the registration experiment, 49 CT and 26 MR images from 20 SIRT patients were selected according to several criteria. The selection procedure is presented in Figure 1. Since the algorithm is designed for registering both CT and MR images in the pre- and post-treatment studies to the ^{99m}Tc -MAA CT, there should be at least one CT and one MR from these studies in addition to the CT image from the ^{99m}Tc -MAA-SPECT/CT for each patient. Since thresholding a ^{99m}Tc -MAA SPECT for lesions may result in overestimation of the tumor targeting performance, landmark delineation (including lesions) for the ^{99m}Tc -MAA study was performed on its CT image. ^{99m}Tc -MAA CTs in our clinical routine are not contrast enhanced. Therefore, it was important to select patients with visible landmarks in ^{99m}Tc -MAA CTs. After patient selection, landmarks visible in all CT and MR images for each patient were chosen and manually delineated by a nuclear medicine physician. Most landmarks were lesions and some of them were vessel knots and metal clips. For evaluation of liver registration, liver contours were manually delineated by a trained researcher and then corrected by a radiographer with over 10 years of experience in liver delineation. For the hyper-parameter tuning of the registration algorithm, 25 CT and 14 MR images from 10 patients were randomly selected for training, 24 CT and 12 MR images from the other 10 patients were used for testing. The characteristics of the 20 SIRT patients are presented in Table 1.

All SIRT datasets for this research were evaluated at KU Leuven after approval by the Ethics Committee Research of UZ / KU Leuven.

B. CNN Structure for Liver Segmentation

The CNN model used in the paper adopts a U-net structure [21], which was developed for biomedical image segmentation. The U-net structure (see Figure 2), adapted for 3D images, consists of four resolution hierarchies with three skip connections to combine the high-level liver features from low resolution hierarchies with the detailed liver features from high resolution hierarchies. Our previous research demonstrated that the CNN segmentations resulted in good segmentation quality without consuming much time and work [23].

C. Segmentation-guided registration

The in-house image-based registration algorithm (see Figure 3) consists of affine registration followed by non-rigid registration. The image similarity metric for affine registration is computed by mutual information. The 12 affine parameters, generated after affine registration, are converted into an initial displacement field for non-rigid registration, where each element represents the displacement of the corresponding voxel in the floating image. The image similarity metric for non-rigid registration is also mutual information. To avoid topology-violating deformations during non-rigid registration, the voxel displacement is regularized through a spring model [24]. Each pair of neighboring voxels is assumed to be connected through a spring, which opposes distance changes. The spring rigidity is used to adjust the regularization power. Therefore, the image-based non-rigid registration algorithm minimizes the weighted sum of the image similarity loss (L_I) and the regularization loss (L_R). The computation and optimization of L_I and L_R are explained in detail in [24].

This image-based registration algorithm was extended through guidance by CNN liver segmentations (see Figure 4). CNN liver segmentations for the reference and floating images are registered via affine transformation to provide initialization of the non-rigid registration. In the next stage, images and their CNN liver segmentations are simultaneously non-rigidly registered. The segmentation similarity loss (L_S) is computed as the sum of squared differences between the reference and warped segmentations, representing the segmentations as binary images. The computation and optimization of L_S are the same as that of L_I when using sum of squared distance as the similarity measurement, as described in [24]. The final loss of the non-rigid registration guided by CNN liver segmentations (CNN-guided) is the weighted sum: $w_I L_I + w_S L_S + L_R$.

To explore the possibility of landmark guidance for better lesion registration, a labeled map containing both CNN liver segmentations and landmarks as two classes replaces the CNN liver segmentations in the CNN-guided algorithm with other settings unchanged. This algorithm guided by both CNN liver segmentations and landmarks (CNN&LM-guided) is compared with the image-based and CNN-guided algorithms in the following experiments.

D. Experiments

CNN liver segmentation. The CNN model trained on CT and MR images from the public challenges and our hospital was implemented to generate automatic liver segmentations for the 49 CT and 26 MR images from 20 SIRT patients selected for the registration experiment. The results were evaluated through the dice similarity coefficient (DSC) between the CNN liver segmentation and the ground truth liver segmentation. The DSC quantifies the overlap between the segmentations, its definition and computation are explained in [22].

Registration of multi-modality images. For affine registration, the results from using either images or CNN liver segmentations were compared. For non-rigid registration, 1D ($w_S = 0$) and 2D grid searches were implemented to find the optimal weights without and with guidance of CNN liver segmentations. The non-rigid registration was initialized by either image-based or CNN-based affine registration. To evaluate the value of landmark guidance for better lesion registration, 2D grid searches were implemented to find the optimal weights for the CNN&LM-guided non-rigid registration initialized by the CNN&LM-based affine registration. Each liver registration was evaluated by the root mean squared distance (RMSD) of

mean surface distance (MSD) between liver contours and mass center distance (MCD) between landmarks: $RMSD = \sqrt{(MSD^2 + \sum_{i=1}^{N_{landmarks}} MCD_i^2)/(1 + N_{landmarks})}$. Mean surface distance is designed to measure the contour difference between two segmentations [25]. Since a landmark can appear with different shapes and volumes in images of different modality, mass center distance, which is independent of shape and volume changes, is used to evaluate landmark registration. The root mean square of RMSDs (RRMSD) from all liver registrations was used to evaluate each pair of weights. The optimal weights, with the lowest RRMSD, were found through grid searches on the training datasets for CT to CT and MR to CT registrations, respectively. After that, different registration settings were compared on the test datasets through RMSD and RRMSD.

Dose estimation. The CNN- and CNN&LM-guided registrations using the optimal weights were implemented to generate registered landmarks (including lesions) for the floating CT and MR images (registered floating landmarks) from the test datasets for comparison with the landmarks delineated on ^{99m}Tc -MAA CTs (reference landmarks). A five-scale Likert score, with its criteria presented in Table 2, was used for grading the floating landmarks registered to the reference landmarks. Dose estimation using the registered floating lesions from the CNN- and CNN&LM-guided algorithms was compared with dose estimation using the reference lesions through the mean dose and the volume percentage receiving at least 70 (V70) and 100 Gy (V100) in the lesion. The absorbed doses of 70 and 100 Gy illustrate an intermediate and high tumor response probability. The injected activities prescribed for the left and right LPTs for each patient were used for dose estimation. LPTs are delineated on CBCTs in our clinical workflow. Dose estimation computed on the left and right LPTs requires registration of CBCT to ^{99m}Tc -MAA-SPECT/CT, which introduces extra potential sources of registration errors. To focus on evaluating the influence of CT and MR registration on dose estimation in this study, the injected activities for the left and right LPTs are summed and distributed in a fractional uptake map generated by normalizing all counts of the ^{99m}Tc -MAA SPECT within the manually delineated liver contour. The computation of the absorbed dose was based on the local deposition model.

III. Results

A. CNN liver segmentation

The DSCs between CNN liver segmentations of CT and MR images from 20 SIRT patients for registration and manual liver segmentations are presented in Figure 5. The median DSCs for CT and MR images are around 0.95 and 0.93, respectively. The DSCs for MR images have an outlier with a very low value of 0.72. Through visual inspection, the CNN segmentation for this MR image misses a large part of the liver volume, and is incapable of registration guidance. Therefore, the case involving this MR image was excluded for MR to CT registration.

B. Registration of multi-modality images

The results of CT to CT registrations are presented in Figure 6, using the optimal weights for each deformable registration. The RRMSD for CNN-based affine registration is 3.1 mm (27%) smaller than that for image-based affine registration. The image-based non-rigid registration increases the RRMSD for CNN-based affine registration by 0.4 mm (5%). The RRMSD for CNN-guided affine and non-rigid registration¹ is 1.0 mm (11%) smaller than that for image-based affine and non-rigid registration. The CNN&LM-guided affine and non-rigid registration² has a RRMSD 2.1 mm (26%) smaller than the CNN-guided one.

The results of MR to CT registrations are presented in Figure 7, using the optimal weights for each deformable registration. The CNN-based affine registration has a RRMSD 3.4 mm (34%) smaller than the image-based one does. The optimal w_I and w_S for CNN-guided non-rigid registration initialized by CNN-based affine registration were both zero. Therefore, the optimal performance of CNN-guided affine and non-rigid registration is the same as that of the CNN-based affine registration. This indicates that both image-based and CNN-guided non-rigid registration degrade the results of CNN-based affine registration. The RRMSD for CNN&LM-guided affine and non-rigid registration is 1.4 mm (21%) smaller than that for CNN-based affine registration.

C. Dose estimation

According to the results in the registration experiment, the optimal performance for CNN&LM- and CNN-guided CT to CT registration and CNN&LM-guided MR to CT registration were achieved when using both affine and non-rigid registrations, while the optimal performance for CNN-guided MR to CT registration was achieved when using only affine transformation. These settings using the optimal weights were used to generate the registered landmarks for dose estimation.

Registration of CT to ^{99m}Tc -MAA CT. The Likert scores for registration of all landmarks (lesions, vessels, clips, etc) are presented in Figure 8a. Over two thirds of registered landmarks had a score of at least 4, which means little or almost no misalignment and insignificant impact on dosimetry. There were in total 29 lesions from 10 SIRT patients used for dose estimation. The relative difference of mean dose and the difference of V70 and V100 between the reference and registered floating lesions using the CNN&LM- and CNN-guided registrations are shown in Figure 9. There were 59% and 45% of

¹ “The CNN-guided affine and non-rigid registration” always refers to the CNN-based affine and CNN-guided non-rigid registration.

² “The CNN&LM-guided affine and non-rigid registration” always refers to the CNN&LM-based affine and CNN&LM-guided non-rigid registration.

lesions having an absolute relative difference of mean dose smaller than 10% for the CNN&LM- and CNN-guided registrations, respectively. Around 79% (76%) and 83% (69%) of lesions have an absolute V70 (V100) difference smaller than 10% for the CNN&LM- and CNN-guided registrations, respectively. As shown in Figure 10, the mean dose, V70, and V100 for the floating lesions registered by both the CNN&LM- and CNN-guided algorithms had a strong correlation ($r \geq 0.95$) with the mean dose, V70, and V100 for the reference lesions.

Registration of MR to ^{99m}Tc -MAA CT. According to the Likert scores for all registered landmarks shown in Figure 8b, around 92% and 63% of lesions registered by the CNN&LM- and CNN-guided algorithms had a score equal to or better than “moderate misalignment and impact on dosimetry”, respectively. There were in total 23 lesions from 10 SIRT patients for dose estimation. The relative difference of mean dose and the difference of V70 and V100 between the reference and registered floating lesions are presented in Figure 11. Around 70% and 43% of lesions had an absolute difference of mean dose smaller than 10 % for the CNN&LM- and CNN-guided registrations, respectively. The CNN&LM- and CNN-guided registrations have around 70% (70%) and 61% (61%) of lesions with an absolute V70 (V100) difference smaller than 10%, respectively. The correlation of mean dose, V70, and V100 between the reference and registered floating lesions is presented in Figure 12. A weaker correlation ($r < 0.90$) of V70 and V100 for the CNN-guided registration is observed than the correlation of mean dose for the CNN&LM- and CNN-guided registrations and of V70 and V100 for the CNN&LM-guided registration.

Some examples of the reference and floating lesions registered by the CNN&LM- and CNN-guided methods and their dose volume histograms (DVHs) are presented in Figure 13.

IV.Discussion

A. Registration of multimodality images

The CNN-based affine registration of CT to CT and MR to CT improves the RRMSD by 3.1 mm (27%) and 3.4 mm (34%), respectively, compared with the image-based affine registration. This substantial decrease shows the advantage of using CNN liver segmentations for affine registration without introducing non-affine deformation. The RRMSD increases by 0.4 mm (5%) when the CNN-guided affine registration is followed by the image-based non-rigid registration for CT images. The CNN-guided non-rigid registration of CT images decreases the RRMSD for the CNN-guided affine registration by 0.1 mm (1%). This indicates the slightly negative influence of using only images (including non-contrast-enhanced ^{99m}Tc -MAA CT) and the limited improvement of using CNN liver segmentations for non-rigid registration of CT images, given the good initialization provided by the CNN-based affine registration. The optimal weights found for the CNN-guided affine and non-rigid MR to CT registration were both zero indicating that using both images and CNN liver segmentations for non-rigid MR to CT registration could not improve the results from the CNN-based affine registration. This might be caused by relatively poorer CNN liver segmentations for MR images than for CT images. Some CNN liver segmentations for MR images were found to miss some low-intensity lesion regions. This can cause big misalignment in these regions due to liver surface matching guided by CNN liver segmentations during the non-rigid registration. Nevertheless, the CNN-guided registration improves the RRMSD by 1.0 mm (11%) and 3.4 mm (34%) for the floating CT and MR images compared with the image-based registration. Even if there might be errors in CNN liver segmentations compared to the “ground truth”, these unedited CNN segmentations are still helpful for improving the image-based registration. This enables the automation of the liver-segmentation-guided registration without the need of extra manual correction for CNN liver segmentations.

Through landmark guidance, the RRMSD is decreased by 2.1 mm (26%) and 1.4 mm (21%) compared with the CNN-guided registration for the floating CT and MR images, respectively. Since we currently don't have an automatic lesion segmentation tool for CT and MR images, manually delineated landmarks were used for both registration guidance and evaluation. This might cause a self-fulfilling effect for registration evaluation. However, it does not change the feasibility of using landmarks for better registration, since manually delineated landmarks can be taken as perfect automatic lesion segmentations. This indicates that developing automatic lesion segmentation would be beneficial for registration guidance. Besides, manually delineated lesions approved by the physician are usable in the clinical context.

B. Dose estimation

A strong correlation ($r > 0.9$) of mean dose estimation existed between the reference and floating lesions registered by the CNN&LM- and CNN-guided registrations. Landmark guidance for the CNN-guided registration resulted in a smaller difference of mean dose for CT to CT registration than for MR to CT registration. Since mean dose is computed on the volume level, it appears less sensitive to contour changes than the voxel-level dosimetry.

For the voxel-level dosimetry, the CNN&LM- and CNN-guided registrations had around 79% (76%) and 83% (69%) of lesions with an absolute V70 (V100) difference between the reference and floating CT lesions smaller than 10%, respectively. A very strong correlation ($r \geq 0.95$) of V70 and V100 existed between the reference and floating CT lesions for the two methods. Landmark guidance for CT to CT registration made small improvement for the voxel-level dosimetry, which was also reflected in the Likert scores given by the physician. Around 70% (70%) and 61% (61%) of MR lesions had an absolute V70 (V100) difference smaller than 10% for the CNN&LM- and CNN-guided registrations, respectively. A weaker correlation ($r < 0.82$) of V70 and V100 was observed between the reference lesions and the floating MR lesions for the CNN-guided registration than for the CNN&LM-guided registration. Landmark guidance for MR to CT registration helped decrease the discrepancy of the voxel-level dose estimation caused by lesion registration.

It was found that the relative difference of mean dose and the difference of V70 and V100 were smaller than 10% for most lesions with a volume over 50 cc. It is reasonable that small lesions with a small shift can create a relatively large voxel

change. Besides, lesions delineated on images of different modality can appear with diverse shape and volume, due to different lesion information expressed in different images or tumor development. Small volumes and large shape and volume differences accounted for a V70 (V100) difference over 10% for 4 (3) out of 6 (7) and 3 (4) out of 5 (9) floating CT lesions registered by the CNN&LM- and CNN-guided algorithms and for 4 (2) out of 7 (7) and 5 (3) out of 9 (9) floating MR lesions registered by the CNN&LM- and CNN-guided algorithms. Good lesion registration does not ensure small difference of dose estimation, since the small size and large shape and volume difference are the other two critical factors with significant impact on dose estimation. It is difficult to eliminate the shape and volume difference between lesions delineated on different images, since each modality reflects a different aspect of lesion appearance. Therefore, it is beneficial to co-register all multi-modality images for joint lesion delineation by the physician, to approach the ground truth delineation by making full use of all information.

Poor lesion registration does not necessarily lead to significant changes of dose estimation. As presented in Figure 13c and 13d, the lesion (green) of radiology MR registered by the CNN-guided method, scored with 2, does not have a good overlap with the reference lesion (red), while both lesion contours include most of the high-uptake region. The relative mean dose difference and the V70 and V100 difference between the reference and registered lesions are -5.2%, -1.0%, and -8.4%, respectively. As long as the reference and registered lesions include a similar area of high- and low-uptake regions, the difference of dose estimation can be insignificant despite of poor registration. This indicates that the volume percentage does not necessarily reflect the true energy deposition for each voxel.

In summary, the performance of the CNN&LM- and CNN-guided registrations makes them useful tools for SIRT treatment planning and verification. The deployment of these semi- and automatic registration tools would allow for dose prediction and measurement based on multi-modality images without introducing much manual interaction and workload, which currently impede the application of image analysis tools in the clinical workflow. The pre- and post-treatment studies contain many images with relatively poor quality, including non-contrast-enhanced CTs from the ^{99m}Tc-MAA study and MR with severe shading or bias artifacts from the ⁹⁰Y-PET/MR. Nevertheless, these registration algorithms can produce reasonably good results for these low-quality images giving them practical value for clinical application. Based on these results, we will study the development of an automatic liver lesion segmentation method for fully automatizing the CNN&LM-guided registration. The clinical influence of these registration methods remains to be fully evaluated in a daily SIRT workflow from a volume-level and voxel-level perspective.

V. Conclusion

Registration guidance using CNN liver segmentations and landmarks greatly improved the performance of the in-house image-based registration. The CNN&LM- and CNN-guided registrations for CT and MR images can be used for the volume-level dosimetry, since the mean doses obtained from the reference and floating lesion contours were very similar. A small V70 and V100 change for most lesions of the floating CT images using the CNN&LM- and CNN-guided registrations demonstrates the feasibility of their application to the voxel-level dosimetry with the physician's checkup. Landmark guidance is needed for the CNN-guided MR to CT registration to be applied to the voxel-level dosimetry. As a result, the CNN&LM- and CNN-guided registration algorithms could become valuable semi- and automatic tools applicable for SIRT dosimetry based on integration of multi-modality information.

VI. Declarations

Funding: This project is funded by the H2020-ITN (MSCA 764458) project Hybrid and by the Research Foundation Flanders (FWO) project G082418N.

Conflicts of interest/Competing interests: Esmael Jafargholi Rangraz is employed by Quirem Medical B.V., Deventer, Netherlands. Christophe M. Deroose is a Senior Clinical Investigator at the Research Foundation Flanders (FWO). Mark Gooding is employed by Mirada Medical Ltd, Oxford, UK, a medical software company. The department of nuclear medicine at KU Leuven receives support from GE for image reconstruction research. No other potential conflicts of interest relevant to this article exist.

Availability of data and material: Due to GDPR, we cannot make the data publicly available.

Code availability: The code used in this study may be made available upon request from the corresponding author.

Authors' contributions: We declare that all authors have contributed to the performed research. Conception and design: XT, EJR, KB, MJG, CMD, JN; Data acquisition: XT, EJR, WC, GM, CV, CMD; Data delineation: XT, RH, WC, CMD; Data analysis: XT, MJG, CMD, JN; Writing manuscript, including contributing to, revising, and approving the manuscript: all.

Ethics approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the Ethics Committee Research of UZ / KU Leuven and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to participate: We obtained the EC approval for use of all patient data involved in this retrospective study.

Consent for publication: All authors of this article give their consent for the publication of identifiable details, which can include figure(s) and/or videos and/or case history and/or details within the text ("Material") to be published in the above Journal.

References

1. Padia SA, Lewandowski RJ, Johnson GE, Sze DY, Ward TJ, Gaba RC, Baerlocher MO, Gates VL, Riaz A, Brown DB, et al. Radioembolization of hepatic malignancies: background, quality improvement guidelines, and future directions. *J Vasc Interv Radiol.* 2017; 28(1):1–15. <https://doi.org/10.1016/j.jvir.2016.09.024>
2. Breedis C, Young G. The blood supply of neoplasms in the liver. *The American journal of pathology.* 1954;30:969–977
3. Bastiaannet R, Kappadath SC, Kunnen B, Braat AJAT, Lam MGEH, de Jong HWAM. The physics of radioembolization. *EJNMMI physics.* 2018;5:22. <https://doi.org/10.1186/s40658-018-0221-z>
4. Ho S, Lau WY, Leung TW, et al. Partition model for estimating radiation doses from Yttrium-90 microspheres in treating hepatic tumours. *European journal of nuclear medicine.* 1996;23:947–952. <https://doi.org/10.1007/bf01084369>
5. De Gersem R, Maleux G, Vanbilloen H, et al. Influence of time delay on the estimated lung shunt fraction on 99mTc-labeled MAA scintigraphy for 90Y microsphere treatment planning. *Clinical nuclear medicine.* 2013;38:940–942. <https://doi.org/10.1097/rnu.0000000000000241>
6. Rangraz EJ, Coudyzer W, Maleux G, Baet K, Deroos CM, Nuyts J. Multi-modal image analysis for semi-automatic segmentation of the total liver and liver arterial perfusion territories for radioembolization. *EJNMMI Res.* 2019; 9(1):19. <https://doi.org/10.1186/s13550-019-0485-x>
7. Lhommel R, van Elmpt L, Goffette P, Den Eynde MV, Jamar F, Pauwels S, Walrand S. Feasibility of 90Y TOF PET-based dosimetry in liver metastasis therapy using SIR-Spheres. *Eur J Nucl Med Mol Imaging.* 2010; 37(9):1654–62. <https://doi.org/10.1007/s00259-010-1470-9>
8. Lee WCC, Tublin ME, Chapman BE. Registration of MR and CT images of the liver: comparison of voxel similarity and surface based registration algorithms. *Comput Methods Programs Biomed.* 2005; 78:101–114. <https://doi.org/10.1016/j.cmpb.2004.12.006>
9. Foruzan AH, Motlagh HR. Multimodality liver registration of Open-MR and CT scans. *Int. J. Comput. Assist. Radiol. Surg.* 2015; 10:1253–1267. <https://doi.org/10.1007/s11548-014-1139-0>
10. Gunay G, Luu MH, Moelker A, van Walsum T, Klein S. Semiautomated registration of pre- and intraoperative CT for image-guided percutaneous liver tumor ablation interventions. *Med Phys.* 2017; 44:3718–3725. <https://doi.org/10.1002/mp.12332>
11. Cazoulat G, Elganainy D, Anderson BM, et al. Vasculature-Driven Biomechanical Deformable Image Registration of Longitudinal Liver Cholangiocarcinoma Computed Tomographic Scans. *Adv. Radiat. Oncol.* 2020; 5:269–278. <https://doi.org/10.1016/j.adro.2019.10.002>
12. Charnoz A, Agnus V, Malandain G, Forest C, Tajine M, Soler L. Liver Registration for the Follow-Up of Hepatic Tumors. *Med Image Comput Comput Assist Interv.* 2005; 8(2):155–162. https://doi.org/10.1007/11566489_20
13. Luu HM, Klink C, Niessen W, Moelker A, van Walsum T. Non-Rigid Registration of Liver CT Images for CT-Guided Ablation of Liver Tumors. *PLoS One.* 2016; 11(9):e0161600. <https://doi.org/10.1371/journal.pone.0161600>
14. Huang X, Wang B, Liu R, Wang X, Wu Z. CT-MR image registration in liver treatment by maximization of mutual information. 2008 IEEE International Symposium on IT in Medicine and Education. 2008; 715–718. <https://doi.org/10.1109/ITME.2008.4743959>
15. Rieder C, Wirtz S, Strehlow J, et al. Automatic alignment of pre- and post-interventional liver CT images for assessment of radiofrequency ablation. *Proc. SPIE.* 2012; 8316: 83163E. <https://doi.org/10.1117/12.911188>
16. Hasenstab KA, Cunha GM, Higaki A, et al. Fully automated convolutional neural network-based affine algorithm improves liver registration and lesion co-localization on hepatobiliary phase T1-weighted MR images. *Eur Radiol Exp.* 2019; 3(1):43. <https://doi.org/10.1186/s41747-019-0120-7>
17. Alsultan AA, van Roekel C, Barentsz MW, et al. The Efficacy of Coil Embolization to Obtain Intrahepatic Redistribution in Radioembolization: Qualitative and Quantitative Analyses. *Cardiovasc Intervent Radiol.* 2020; 43:391–401. <https://doi.org/10.1007/s00270-019-02351-1>
18. Spahr N, Thoduka S, Abolmaali N, Kikinis R, Schenk A. Multimodal image registration for liver radioembolization planning and patient assessment. *Int J Comput Assist Radiol Surg.* 2019; 14(2):215–225. <https://doi.org/10.1007/s11548-018-1877-5>
19. Nodari G, Popoff R, Riedinger JM. Impact of contouring methods on pre-treatment and post-treatment dosimetry for the prediction of tumor control and survival in HCC patients treated with selective internal radiation therapy. *EJNMMI Res.* 2021; 11(1):24. <https://doi.org/10.1186/s13550-021-00766-x>
20. Jafargholi Rangraz E, Tang X, Van Laeken C, et al. Quantitative comparison of pre-treatment predictive and post-treatment measured dosimetry for selective internal radiation therapy using cone-beam CT for tumor and liver perfusion territory definition. *EJNMMI Res.* 2020; 10(1):94. <https://doi.org/10.1186/s13550-020-00675-5>
21. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. <https://arxiv.org/abs/1505.04597>
22. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad Radiol.* 2004; 11(2):178–189. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8)
23. Tang X, Jafargholi Rangraz E, Coudyzer W, et al. Whole liver segmentation based on deep learning and manual adjustment for clinical use in SIRT. *Eur J Nucl Med Mol Imaging.* 2020; 47, 2742–2752. <https://doi.org/10.1007/s00259-020-04800-3>
24. De Moor K, Nuyts J, Plessers L, Stroobants S, Maes F, Dupont P. Non-rigid registration with position dependent rigidity for whole body PET follow-up studies. In: 2006 IEEE Nuclear Science Symposium Conference Record, vol. 6: 2006. p. 3502–3506. <https://doi.org/10.1109/nssmic.2006.353755>

25. Heimann T, van Ginneken B, Styner sM, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging. 2009;28(8):1251–1265. <https://doi.org/10.1109/tmi.2009.2013851>

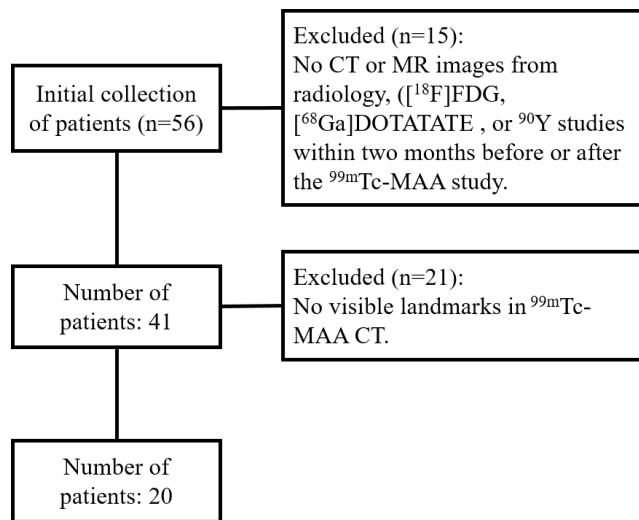


Fig.1 SIRT patient selection for the registration experiment

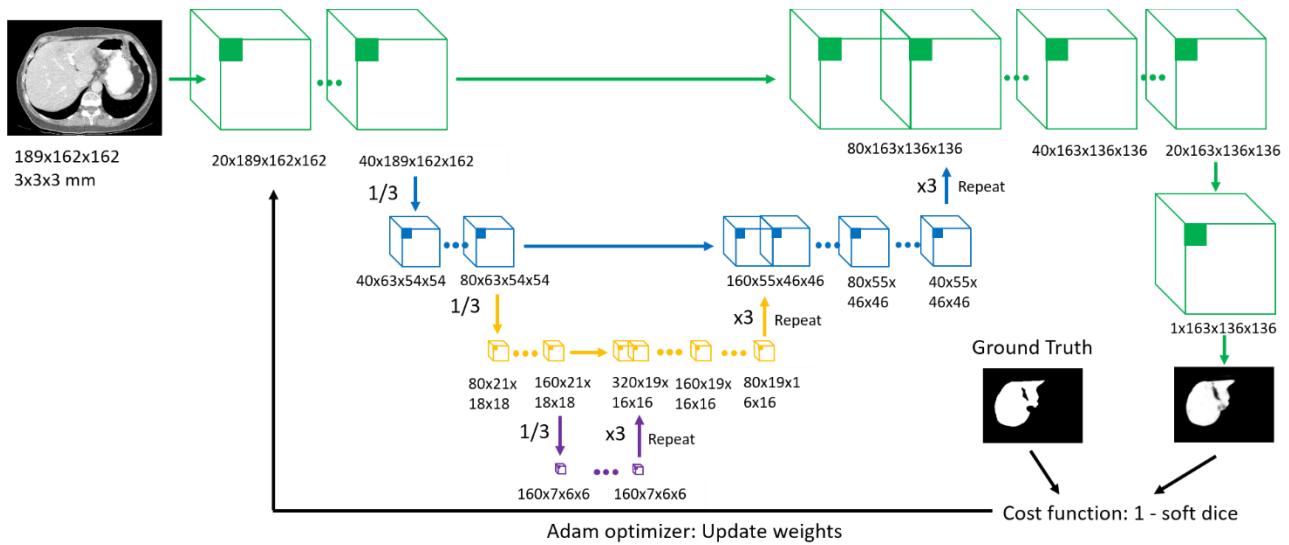


Fig.2 Overview of the CNN structure. The model consists of four resolution hierarchies with three skip connections. The input image has the voxel size of $3 \times 3 \times 3 \text{ mm}^3$. The input image size is $189 \times 162 \times 162$ and the output image size is $163 \times 136 \times 136$. The up- and down-sampling rate is three and the up-sampling is implemented by repetition. The convolutional kernel size for all layers is $3 \times 3 \times 3$

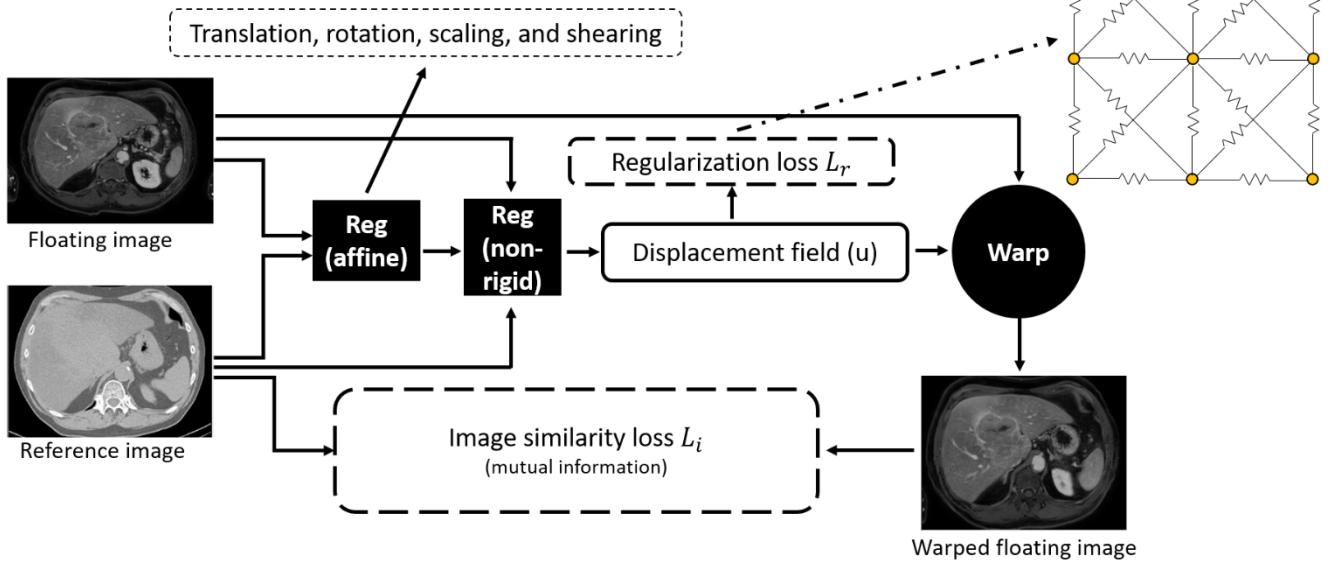


Fig.3 Overview of the in-house image-based registration algorithm

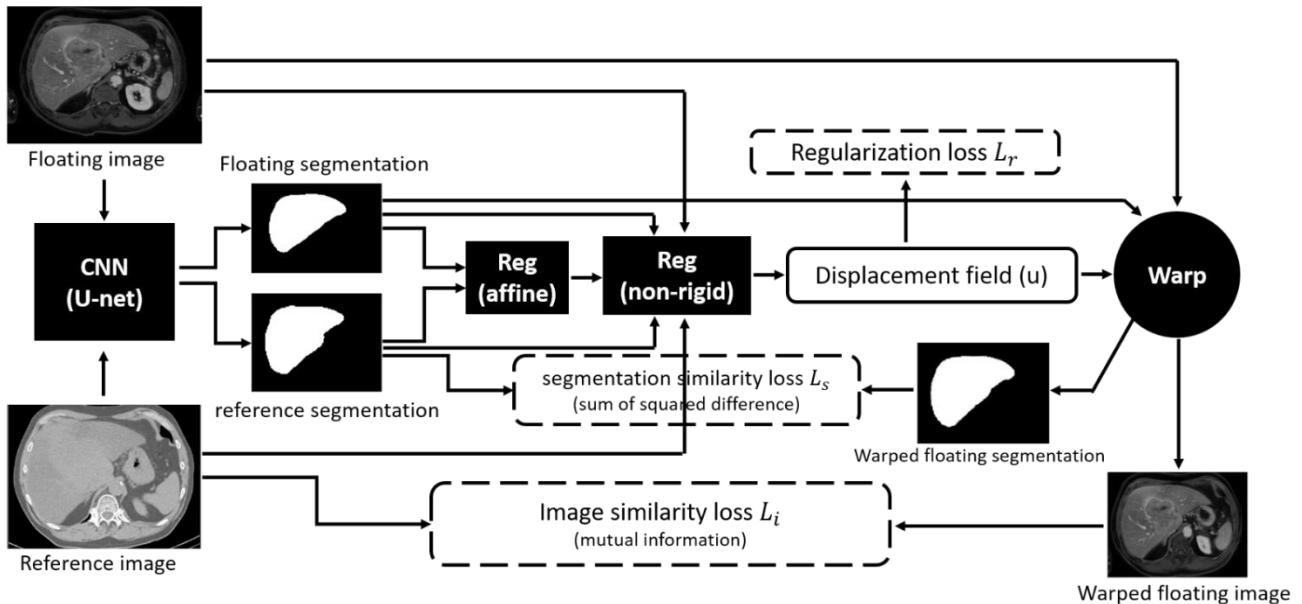


Fig.4 Overview of the registration algorithm guided by CNN liver segmentations

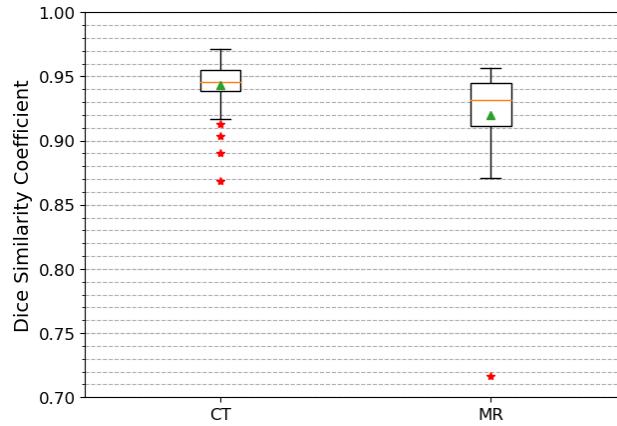


Fig.5 DSCs between manual liver segmentations and CNN liver segmentations of 49 CT and 26 MR images from the 20 SIRT patients selected for the registration experiment. The orange line, green triangle, and red stars represent median, mean, and outliers, respectively. The box corresponds to the first and third quartiles and the whiskers give the range (except for the outliers).

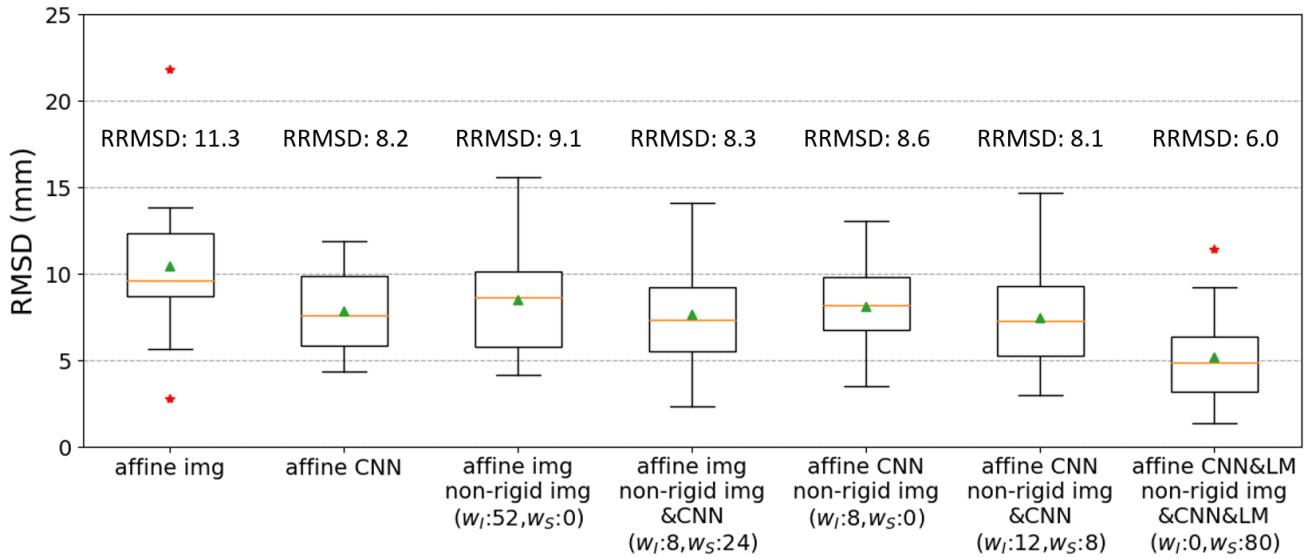


Fig.6 RMSDs and RRMSDs of the test datasets (14 floating CT images) for comparison of image-based, CNN-guided, and CNN&LM-guided CT to CT registrations

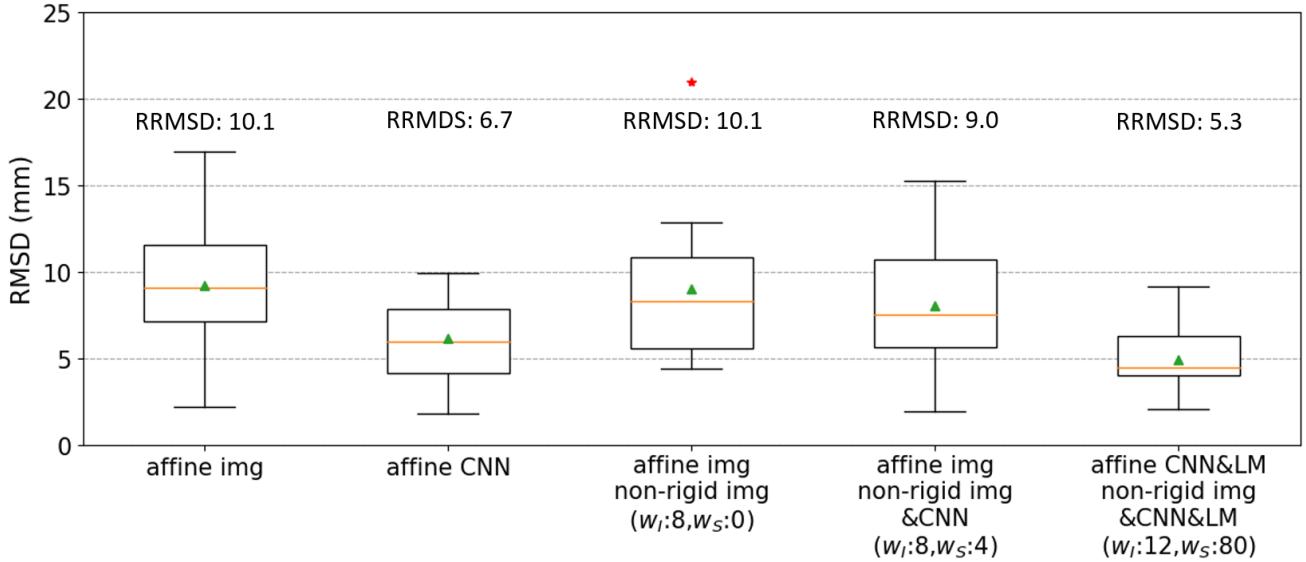


Fig.7 RMSDs and RRMSDs of the test datasets (12 floating MR images) for comparison of image-based, CNN-guided, and CNN&LM-guided MR to CT registrations

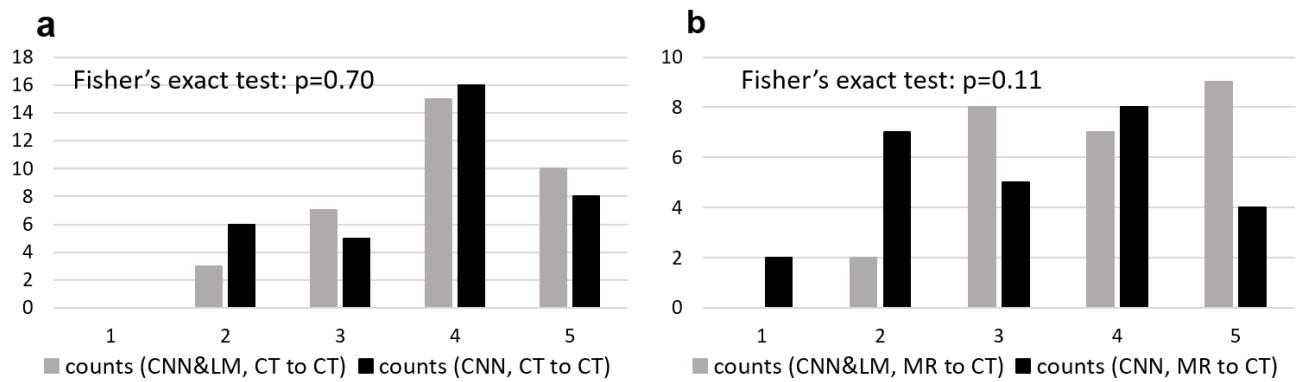


Fig.8 Likert scores for the registered landmarks of floating CT (a) and MR (b) images using the CNN&LM-guided and CNN-guided registrations

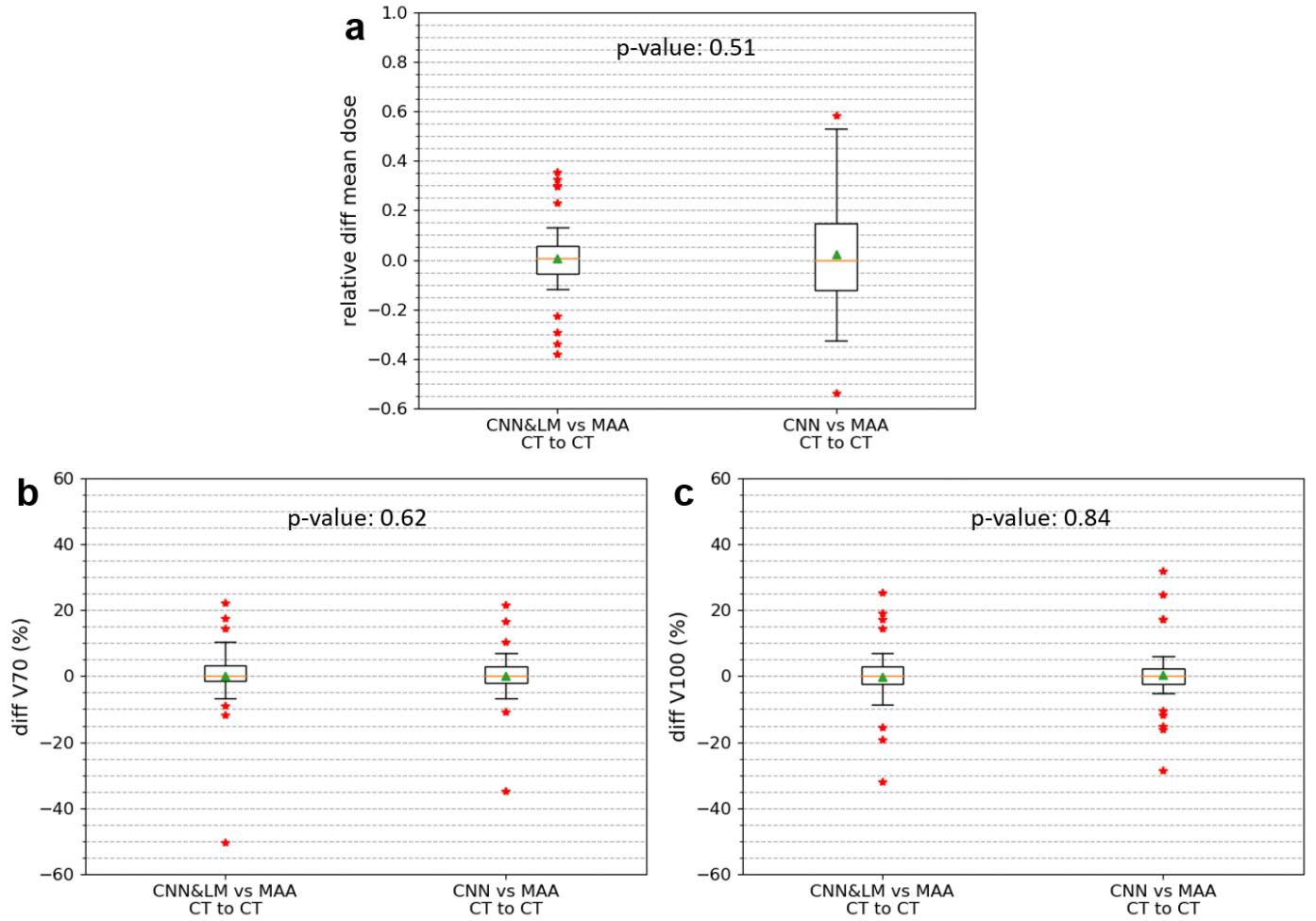


Fig.9 Relative difference of mean dose (a) and difference of V70 (b) and V100 (c) between the reference and registered floating lesions using the CNN&LM-guided or CNN-guided CT to CT registration. MAA: the reference lesions. CNN&LM: the floating lesions registered by the CNN&LM-guided registration. CNN: the floating lesions registered by the CNN-guided registration. The relative difference of mean dose is computed by $(\text{mean dose (floating)} - \text{mean dose (reference)}) / \text{mean dose (reference)}$. The difference of V70 and V100 is computed by $(V70 \text{ or } V100 \text{ (floating)} - V70 \text{ or } V100 \text{ (reference)})$

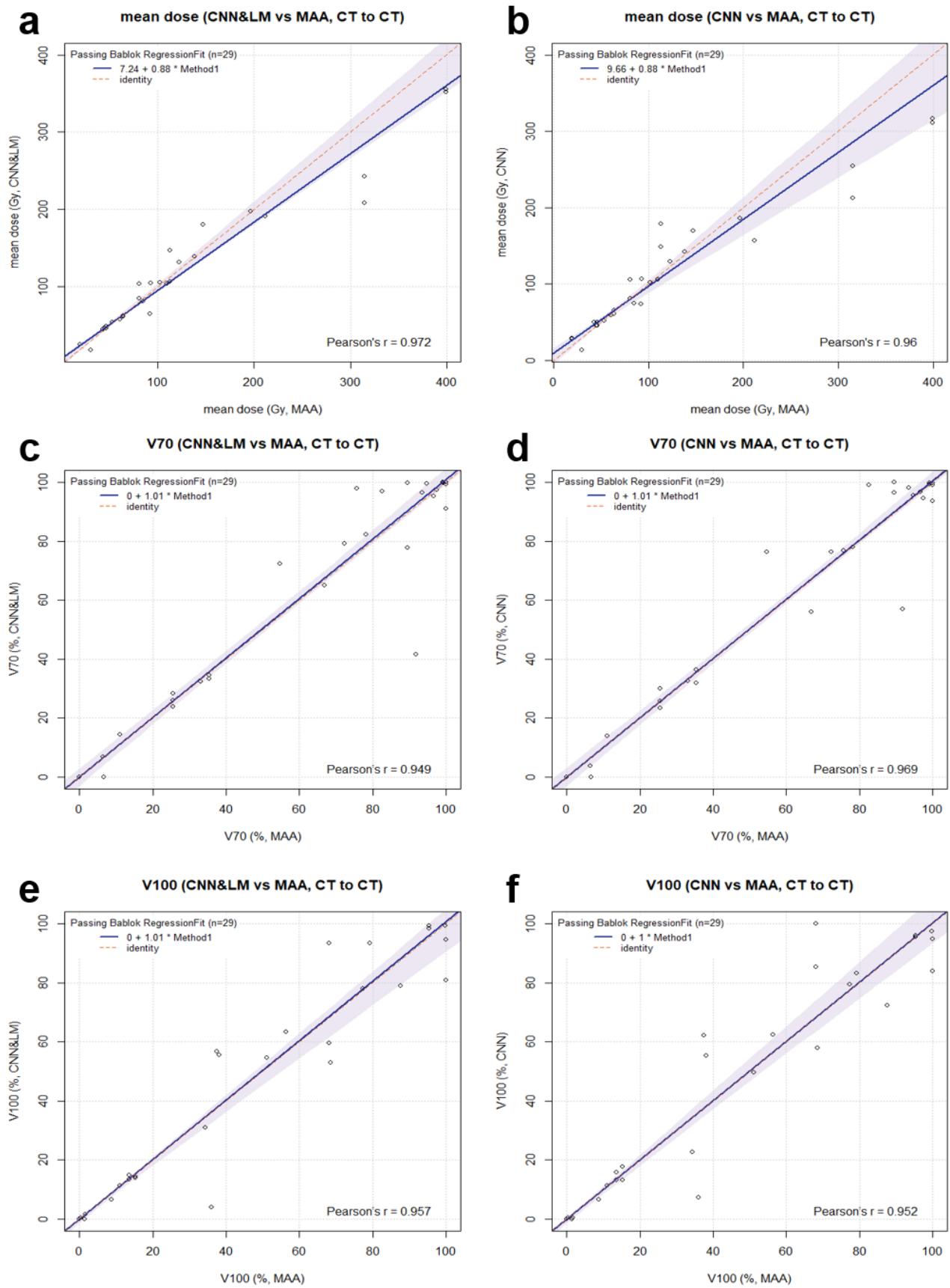


Fig.10 Passing-Bablok plots for the mean dose, V70, and V100 estimated on the floating CT lesions registered by the CNN&LM- or CNN-guided algorithm versus the mean dose, V70, and V100 estimated on the reference lesions. MAA: the reference lesions. CNN&LM: the floating lesions registered by the CNN&LM-guided registration. CNN: the floating lesions registered by the CNN-guided registration

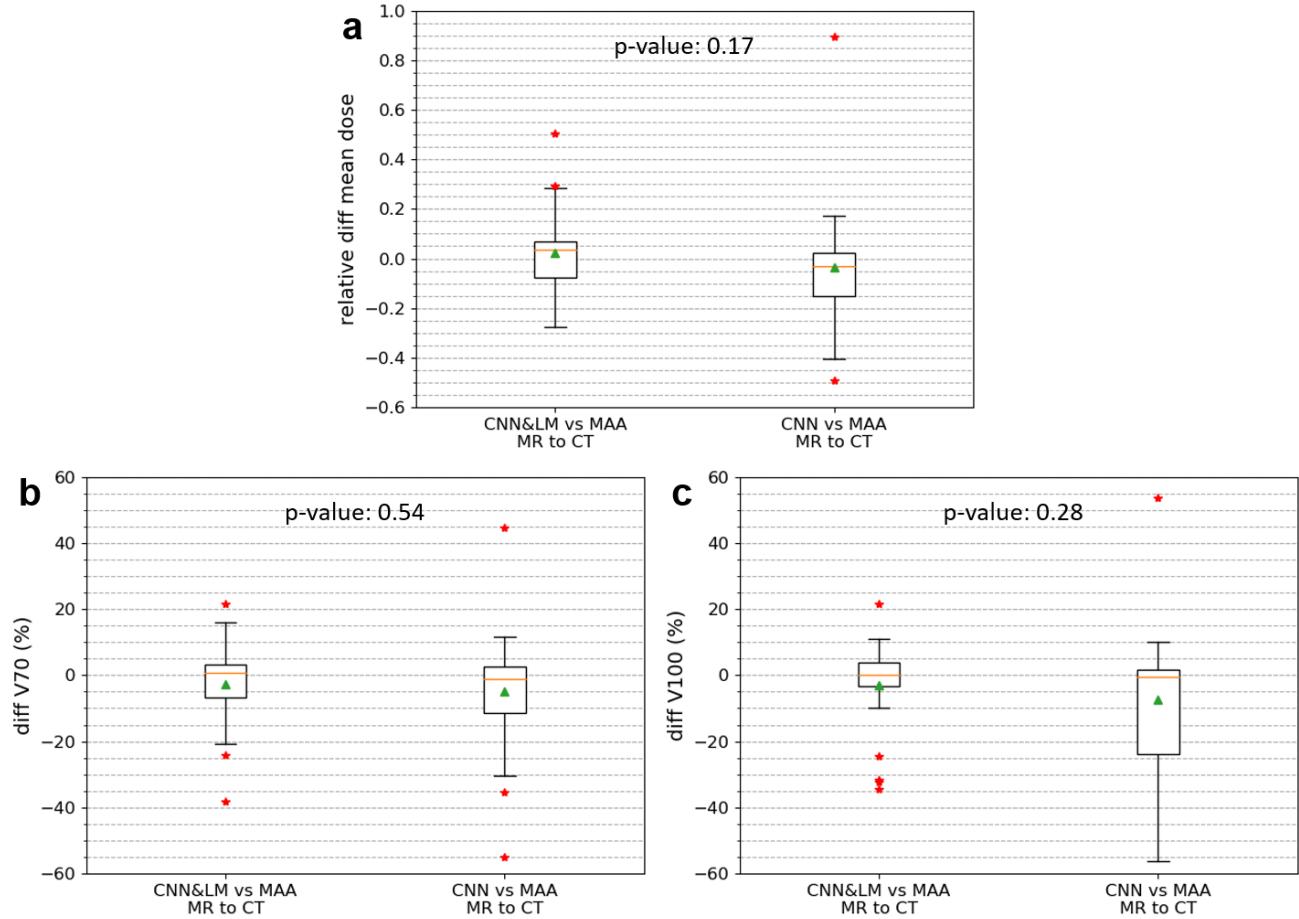


Fig.11 Relative difference of mean dose (a) and difference of V70 (b) and V100 (c) between the reference and registered floating lesions using the CNN&LM-guided or CNN-guided MR to CT registration. MAA: the reference lesions. CNN&LM: the floating lesions registered by the CNN&LM-guided registration. CNN: the floating lesions registered by the CNN-guided registration. The relative difference of mean dose is computed by $(\text{mean dose (floating)} - \text{mean dose (reference)}) / \text{mean dose (reference)}$). The difference of V70 and V100 is computed by $(V70 \text{ or } V100 \text{ (floating)} - V70 \text{ or } V100 \text{ (reference)})$

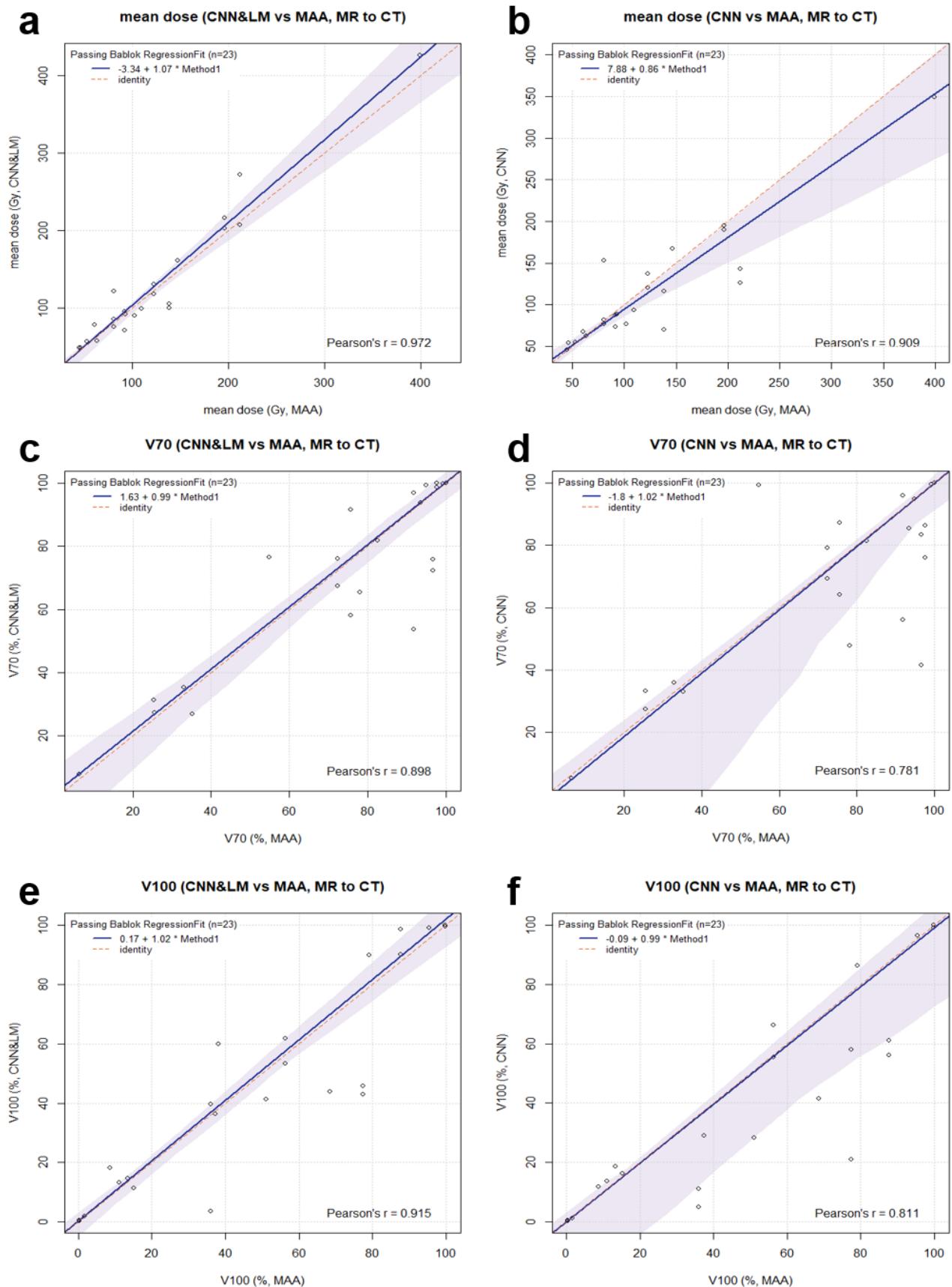


Fig.12 Passing-Bablok plots for the mean dose, V70, and V100 estimated on the floating MR lesions registered by the CNN&LM- or CNN-guided algorithm versus the mean dose, V70, and V100 estimated on the reference lesions. MAA: the reference lesions. CNN&LM: the floating lesions registered by the CNN&LM-guided registration. CNN: the floating lesions registered by the CNN-guided registration

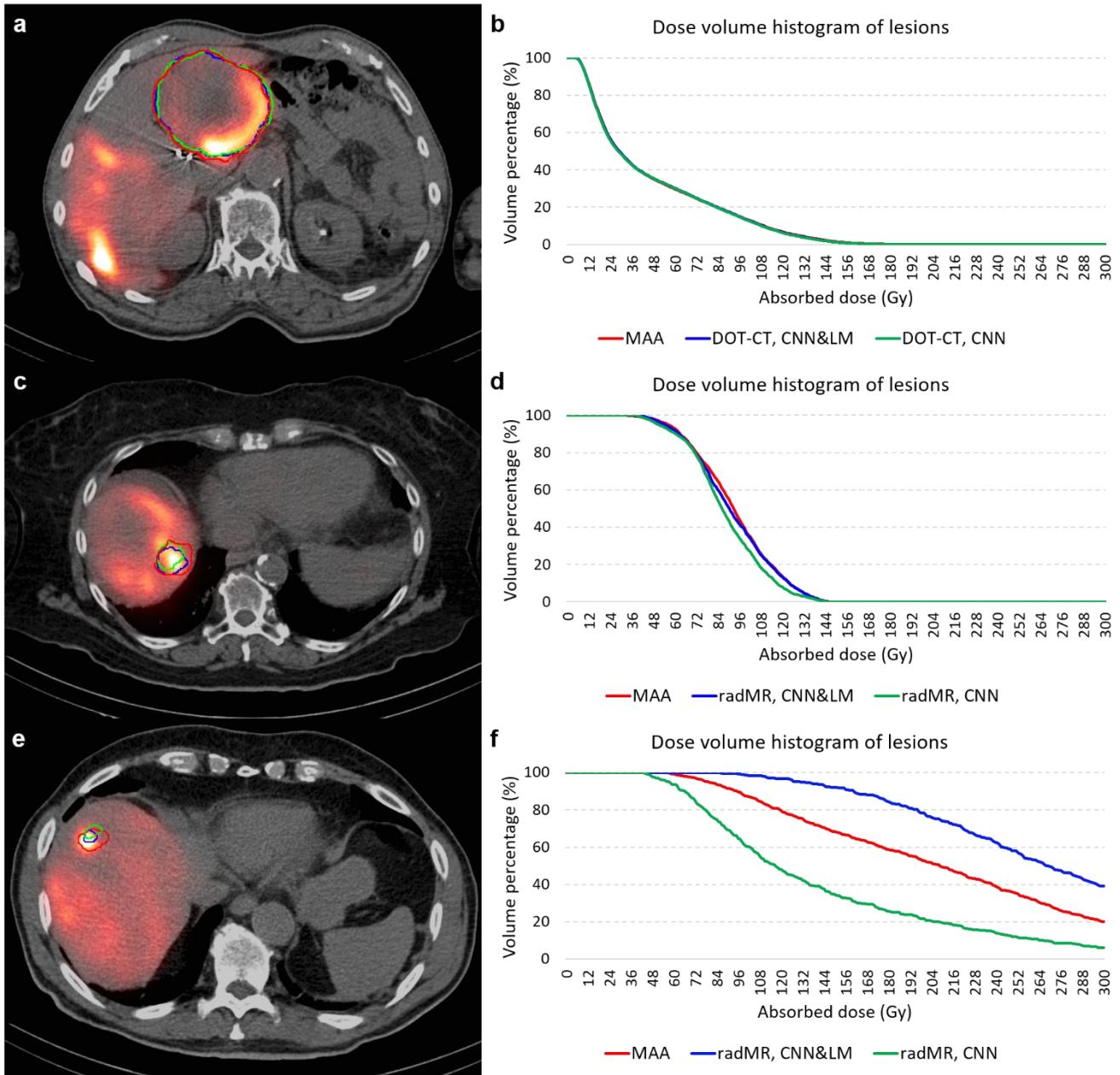


Fig.13 Examples of the reference and floating lesions registered by the CNN&LM- and CNN-guided methods and their dose volume histograms from three patients. The red, blue, and green contours represent the reference lesion (MAA), the floating lesion registered by the CNN&LM-guided method (CNN&LM), and the floating lesion registered by the CNN-guided method (CNN). **a, b** the registered floating lesions from the CT of the DOTATATE (DOT) study; CNN&LM: score is 5, relative difference of mean dose is -0.1%, the differences of V₇₀ and V₁₀₀ are 0.3% and -0.1%, CNN: score is 5, relative difference of mean dose is -0.9%, the differences of V₇₀ and V₁₀₀ are 0.2% and -0.4%. **c, d** the registered floating lesions from the radiology MR (radMR); CNN&LM: score is 3, relative difference of mean dose is -1.3%, the differences of V₇₀ and V₁₀₀ are -0.7% and -1.1%, CNN: score is 2, relative difference of mean dose is -5.2%, the differences of V₇₀ and V₁₀₀ are -1.0% and -8.4%. **e, f** the registered floating lesions from radMR; CNN&LM: score is 3, relative difference of mean dose is 28.4%, the differences of V₇₀ and V₁₀₀ are 2.5% and 11.1%, CNN: score is 2, relative difference of mean dose is -32.5%, the differences of V₇₀ and V₁₀₀ are -11.1% and -26.5%

TABLE 1: Characteristics of the SIRT patients for registration experiment

Characteristics	Training	Test
No. of SIRT patients	10	10
Age (y), median [range]	66.5 [46, 75]	64.5 [25, 78]
Sex (female/male)	5/5	4/6
Weight (kg), median [range]	71 [46,105]	75 [46,95]
Height (m), median [range]	1.70 [1.55, 1.74]	1.70 [1.54, 1.78]

TABLE 2: Likert score criteria for scoring landmark registration

Score	Criteria
1	Major misalignment exists for the registration. Major impact on dosimetry is expected. Dosimetry results are deemed unreliable.
2	Pronounced misalignment exists for the registration. Substantial impact on dosimetry is expected.
3	Moderate misalignment exists for the registration. Moderate impact on dosimetry is expected.
4	Little misalignment exists for the registration. No significant impact on dosimetry is expected.
5	Near perfect alignment for the registration. No intervention is warranted and dosimetry is deemed reliable.