Journal of Medical Systems (2019) 43: 194
https://doi.org/10.1007/s10916-019-1256-y

IMAGE & SIGNAL PROCESSING

A supervised network for fast image-guided radiotherapy (IGRT) registration

Zhixin Yao [1,2] · Hansheng Feng [1,2] · Yuntao Song [1,2] · Shi Li [1,2] · Yang Yang [1] · Lingling Liu [3,4] · Chunbo Liu [2]

Received: 15 November 2018 / Accepted: 27 March 2019 / Published online: 22 May 2019

Abstract

3D/3D image registration in IGRT, which aligns planning Computed Tomography (CT) image set with on board Cone Beam CT (CBCT) image set in a short time with high accuracy, is still a challenge due to its high computational cost and complex anatomical structure of medical image. In order to overcome these difficulties, a new method is proposed which contains a coarse registration and a fine registration. For the coarse registration, a supervised regression convolutional neural networks (CNNs) is used to optimize the spatial variation by minimizing the loss when combine the CT images with the CBCT images. For the fine registration, intensity-based image registration is used to calculate the accurate spatial difference of the input image pairs. A coarse registration can get a rough result with a wide capture range in less than 0.5 s. Sequentially a fine registration can get accurate results in a reasonable short time. RSD-111 T chest phantom was used to test our new method. The set-up error was calculated in less than 10s in time scale, and was reduced to sub-millimeter level in spatial scale. The average residual errors in translation and rotation are within ±0.5 mm and ± 0.2°.

Keywords IGRT · CBCT · CNNs · Image registration · Intensity-based registration

Introduction

Radiation therapy is a widely used, highly effective treatment method for cancers. Among the numerous methodologies and techniques that have been developed, a very important improvement for radiation therapy is the image-guided radiation therapy (IGRT). IGRT is a booming technique which can correct the set-up error of cancer patient to improve treatment accuracy [1–5]. The on board CBCT-based IGRT is becoming a popular method for irradiating the tumor and inducing as little as possible damage to the surrounding tissues. CBCT which is assembled on the gantry of a proton therapy system establishes a coordinate system relative to the treatment room. Meanwhile a planning CT determines the exact position of a tumor. By utilizing a registration algorithm, the planning CT images are matched with the CBCT images. As a sequence, the spatial difference is calculated, in order to correct the set-up error of the patients.

The IGRT registration is dominated by bone and not soft tissue [3]. It usually uses the automated rigid registration method, which has two categories: intensity-based registration and feature-based registration. The intensity-based image registration is a global exhaustive search strategy for automated rigid registration. The robustness of intensity-based 3D image registration relies on multi-resolution strategy with local optimizers. The intensity-based registration promises high accuracy, however it is quite time-consuming [6]. In order to reduce the computational complexity, many optimization strategies have been developed, e.g. simulated annealing [7] and genetic algorithm [8]. However by far the computational cost is still high. On the other hand, the feature-based image registration is a semi-global search strategy. Relative to the

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Hansheng Feng
hsfeng@ipp.ac.cn

[1] Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

[2] University of Science and Technology of China, Hefei 230026, China

[3] Cancer Hospital, Chinese Academy of Science, Hefei 230031, China

[4] Anhui Province Key Laboratory of Medical Physics and Technology, Center of Medical Physics and Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

intensity-based method, the computational cost is greatly reduced. For the feature-based image registration, the images must be segmented [9]. Anatomical features are required for matching a pair of images. Therefore the accuracy is affected by the segmentation methods.

Recently there are increasing interests in using CNNs in medical image registration [10]. CNNs is a feed-forward network and it is considered to be most suitable for image processing tasks [11]. It is suggested that a neural network model can quickly find the approximate pose of 3D objects and align them without solving an iterative optimization. CNNs performs well in image segmentation, image recognition, and image classification. By far there are only a few researches who use CNNs for image registration. Liao, R., et al. [12] proposed a reinforcement learning algorithm for 3D global rigid registration. The algorithm improves the accuracy and robustness compared to state-of-art image registration, however the computation is still relatively slow. Salehi, S.S.M., et al. [13] proposed real-time deep registration for fetal MRI data, The computation speed is improved, but the output accuracy needs to be further optimized.

In this work, a hierarchical registration framework which combines the conventional method and regression CNNs is proposed. Our new method takes advantages of the regression CNNs as well as the conventional method to simultaneously optimize both the speed and accuracy for IGRT.

Materials and methods

In image registration the goal is to determine the coordinate transform $T: I_{floating} \rightarrow I_{fixed}$ which aligns the floating image $I_{floating}$ and the fixed image I_{fixed} at the same positions, and to determine the coordinate transform $T(pitch, yaw, roll, t_X, t_Y, t_Z)$, which is an important parameter for adjusting the treatment couch to correct the set-up error of the patients.

The schematic diagram of our new registration method is shown in fig. 1.The left part in red box shows the framework of the CNNs. A pair of image datasets are used to train the neural network model, which gives out an initialized transform. This initialized transform connects the CNNs with the intensity-based registration. On the right side in green box is a conventional intensity-based registration, which requires a fixed image, a floating image and an initialized transform. The CNNs works as the coarse registration, it used the neural network model to extract the 3D features of the input image pairs and calculated the spatial variation. The neural network model gives out the rough result initialized transform in less than 0.5 s. The initialized transform smooths the spatial variation between the fixed image and the floating image, and it reduces the capture range of intensity-based registration. The intensity-based registration works as the fine registration. Due to the initialized transform, the number of optimization iterations for intensity-based registration is reduced. The search range of the optimizer is decreased, a critical drawback that the cost functions associated with intensity similarity metrics are often con-convex is overcome. Therefore the fine registration can give out a more accurate result with fewer optimization iterations. This new method can shorten the calculation time of image registration while ensure the accuracy.

The method of 3D-3D registration network is shown in fig. 2. It includes (1) input, which contains a pair of images with label. The label is the rigid transformation matrix between a fixed image and a floating image, which is employed as the ground truth for training the neural network model. (2) 3D feature extraction, 3D convolution layers and pooling layers are used to extract the feature of an image pair. (3) The regression which is used to predict the spatial variation. The variation is defined by the translation and rotation matrix. The input, 3D feature extraction, and regression are explained in details in following sections:

Input

Each dataset (floating or fixed) contains 100 images. The CBCT images were acquired on Elekta Synergy X-ray volume imaging (XVI), and the CT images were acquired on Philips Briliancei CT from patients at different scene: head, abdomen,

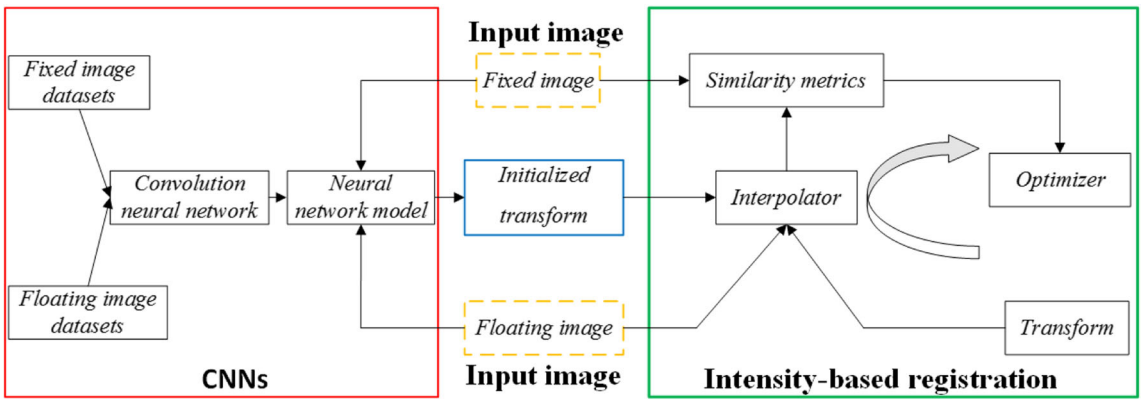

Fig. 1 Schematic diagram of the proposed registration method

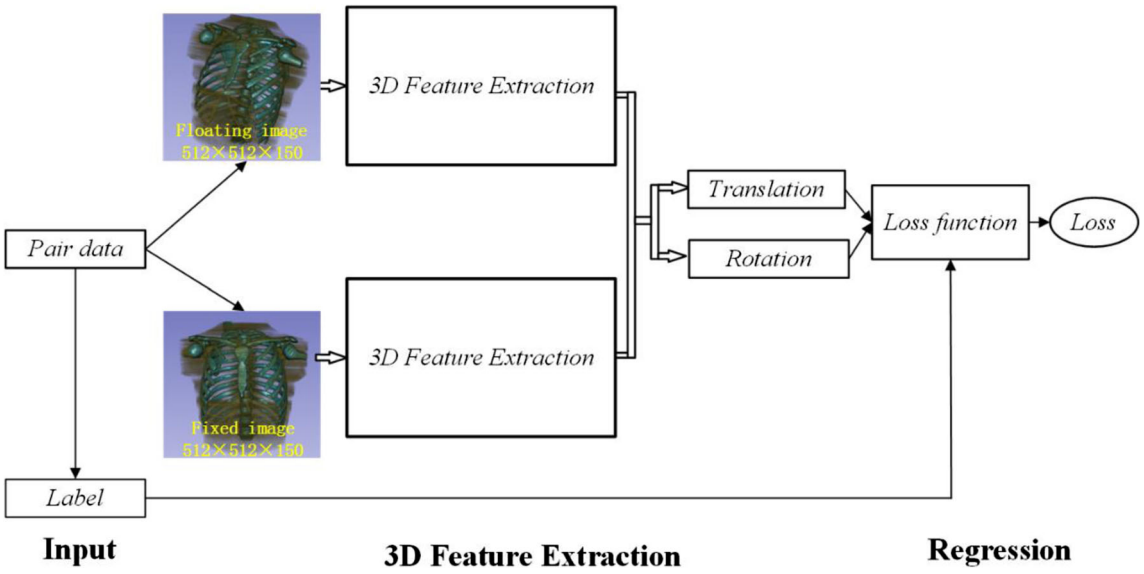

Fig. 2 The framework of the network

chest, and pelvic (25 pairs of images at each part). Each scene includes the most common location of the tumors. For example, "Head" includes thyroid tumor, laryngeal cancer and nasopharynx cancer; "Abdomen" includes gastric cancer, liver cancer and pancreatic tumor; "Chest" includes lung cancer and esophagus cancer; "Pelvic" includes bladder tumor and prostate tumors. By an oncologist, each pair of the CT and CBCT images was aligned using a same coordinate, so we assumed that there was no spatial variation between them.

The CT and CBCT images have different field of views (FOVs). The scene is illustrated in the Fig. 3, where the CT image is shown in gray, and the CBCT image is shown in color (in yellow square). Because the CT images have larger FOVs, we cropped the CT images and resized each image to $512 \times 512 \times 150$ with voxel spacing 1.0 mm × 1.0 mm × 2.0 mm to fit the CBCT images.

The processed CBCT images and CT images share the same size, and location. An affine matrix, which is the exacted spatial variation between the fixed image (CT) and the floating image (CBCT) was used to transform the CBCT images. The affine was defined as the label, which is the supervised value of the network.

3D feature extraction architecture

The framework of a 3D feature extraction is shown in Fig. 4. The network consisted of 5 3D convolutional layers followed by 3 fully connected layers. Initially, 3D convolutional layers and 3D max-pooling layers were alternatively stacked to extract 3D feature of input images. The convolutional layers used 8, 32, 32, 64, 64 filters, with $2 \times 2 \times 2$ kernels. ReLU nonlinear function and batch normalization were employed after each convolutional layer. The first and second convolutions were followed by two $2 \times 2 \times 2$ max pooling layers. After the Conv5 layer the input image was converted into 64 3D features. Next these 3D feature vectors were fed into the fully-connected layers. Each fully-connected layer has 512,512,256 activation neurons. Finally, the fully-connected layers integrated these 3D features into a row vector, whose size is 256. The row vectors of the inputted volumes were imported into the CNN regression to predict the differences between the inputted volumes, each predicted value has 6 elements corresponding to the transformation matrix($pitch$, yaw, $roll$, t_X, t_Y, t_Z).

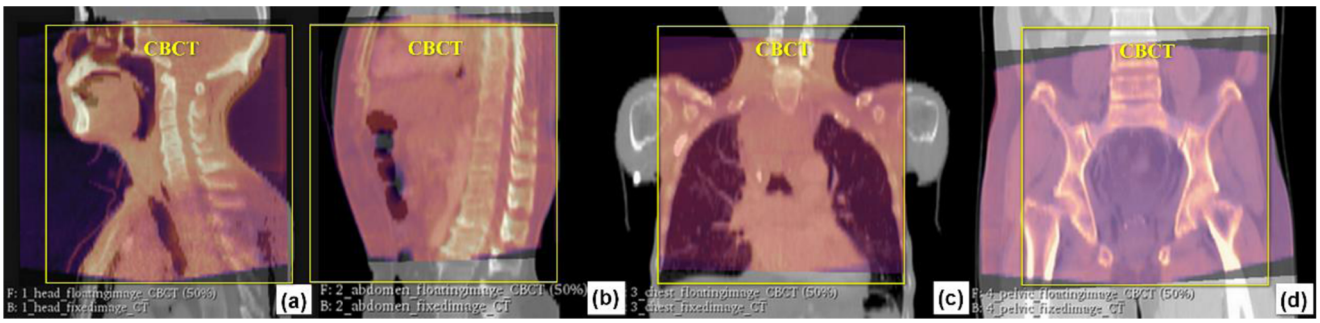

Fig. 3 Examples of the image sets for a) head; b) abdomen; c) chest; d) pelvic. These four pairs of images were from different patients

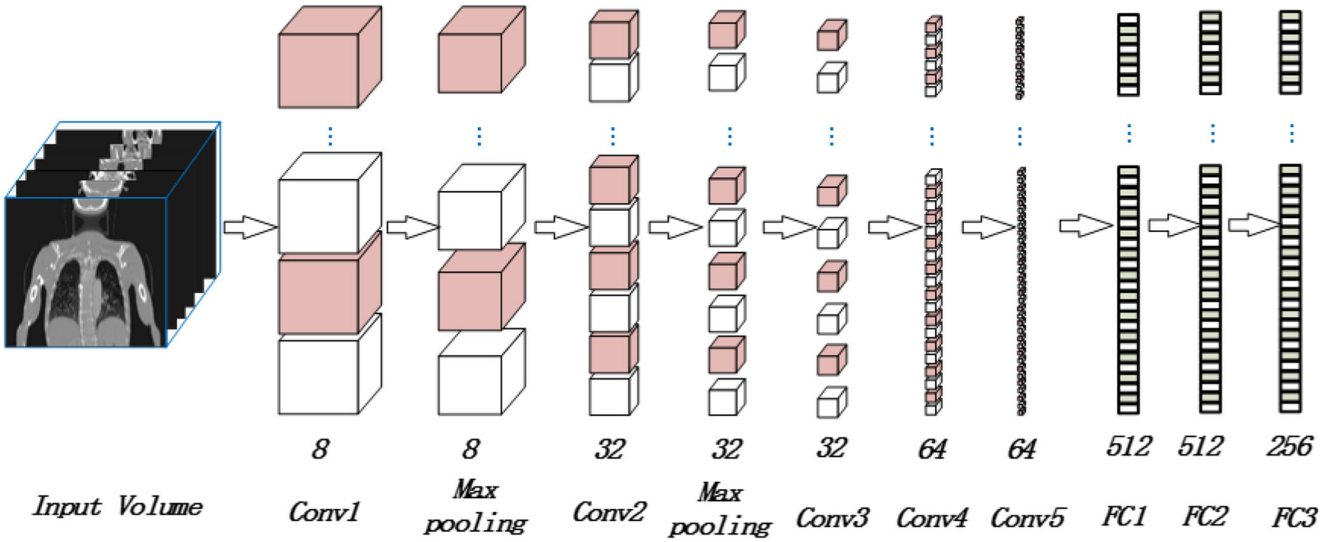

Fig. 4 The framework of 3D feature extraction

Regression

After the 3D feature extraction, the CNN regression predicted the difference between a fixed image and a floating image [14, 15], which was expressed by a translation matrix and a rotation matrix. The label is the supervised value of network and the exact difference of image pairs. The loss function is the difference between the predicted value and the label. Training the networks involved iterations of back-propagation with the total loss function. In order to improve the performance of the loss function, the affine matrix is shortened to six parameters.

$$T(pitch, yaw, roll, t_X, t_Y, t_Z) = \begin{bmatrix} 1 & 0 & 0 & t_X \\ 0 & \cos(pitch) & -\sin(pitch) & t_Y \\ 0 & \sin(pitch) & \cos(pitch) & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos(yaw) & 0 & \sin(yaw) & 0 \\ 0 & 0 & 0 & 0 \\ -\sin(yaw) & 0 & \cos(yaw) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos(roll) & -\sin(roll) & 0 & 0 \\ \sin(roll) & \cos(roll) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The loss function expresses the distance between the prediction and the label. The total loss consists of the translation loss and the rotation loss.

$$Loss_{Total} = Loss_{Rotation} + Loss_{Translation} \quad (2)$$

It is difficult to balance the translation loss and the rotation loss because the units are different. Therefore, target registration error (TRE) [16] was used as the loss function. TRE is defined as the distance between the fixed image and the floating image.

$$TRE = \frac{1}{N}\sum_{i=1}^{N} \| T_i(pitch, yaw, roll, t_X, t_Y, t_Z) - T_0(pitch, yaw, roll, t_X, t_Y, t_Z) \| \quad (3)$$

Where N is the number of experiments. T_i is the calculated result, and the T_0 is the ground truth.

When compared with the image registration methods, researchers often use TRE to measure the accuracy. We used TRE as the loss function, as it provided an exact difference between the predicted value and the label. During every back-propagation, the network provided an accurate result, so the neural network model could be trained more efficiently and give out results with good accuracy.

Results

Datasets

Training dataset

25 pairs of images from each part (head, chest, abdomen and pelvic) of patients were used as golden standard to train the neural network model. As discussed in the previous sections, there was no spatial variation between each pair of images. The CBCT images were rotated and translated along the X, Y, and Z axes between −5° to +5° and between −30 to +30 mm, respectively. During the transforming processes, linear interpolation was employed in order to resample the CBCT images. The transformed CBCT images and the corresponding CT images were used to determine the set-up error of the patient during radiotherapy. The total number of the training data are 20,000 volumes. One quarter each for head, chest, abdomen and pelvic images. The distribution of transformed matrix are shown in the Fig. 5. It can be found that the spatial variation of the head image sets covers every angle for rotation and position for translation within the capture range. Images from the other three parts of patients have similar distributions that the floating image were rotated and tranlated randomly.

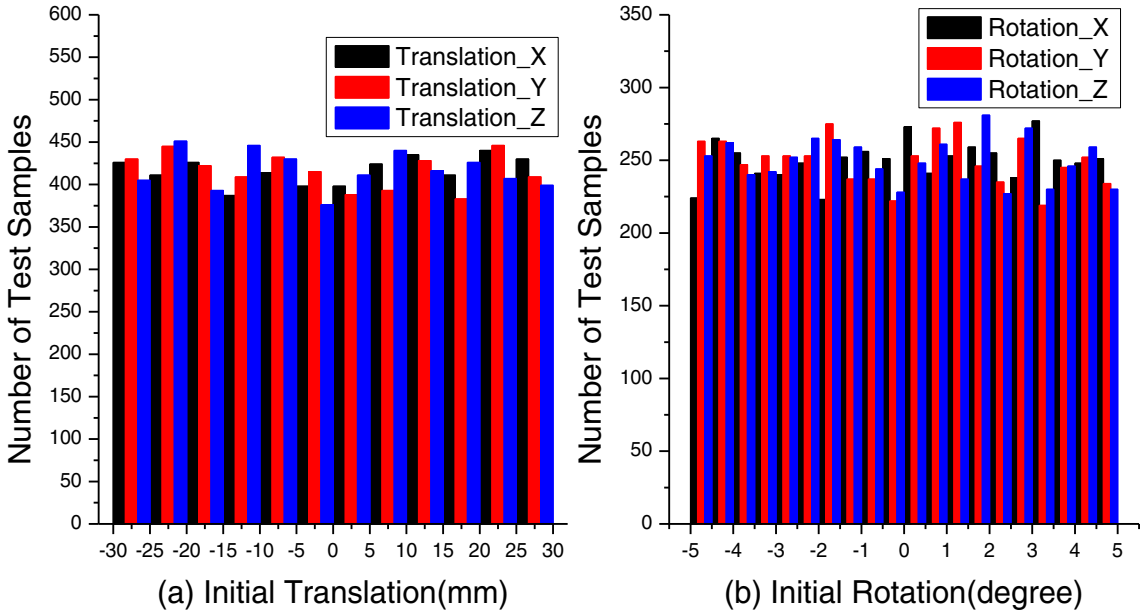

Fig. 5 Histogram of (a) initial rotation and (b) translation of the training datasets. The initial transform and rotation were distributed uniformly to make sure that all situations were considered

Testing dataset

In order to test the accuracy and generalization of the neural network model, 10 images pairs from each part of the patient were analyzed. In total, 40 pairs of images were rotated and translated randomly. 100 affine matrices were applied to the floating images. The affine matrices were distributed in the capture range that fully covered the angles between −5° to +5° and positions between −30 to +30 mm. There were 4000 volumes used for test samples in total.

Result

An RSD-111 T chest phantom was used to evaluate 3 different algorithms: Intensity-based registration (IBR), CNNs registration (CNNR) and our new CNNs and intensity-based registration (CIR). A pair of images of RSD-111 T chest phantom were used to test each method. The CT images were acquired using Philips Briliancei CT, and the CBCT images were acquired by the Elekta Synergy XVI. The physical size of the CBCT and CT images is 512 × 512 × 150 and the voxel spacing is 1.0 mm × 1.0 mm × 2.0 mm.

High quality CBCT and CT images show clearly the internal structures and the edges of bones and lungs [17]. They were acquired under a same coordinate, and the oncologist adjusted precisely to align them exactly. One example is shown in fig. 6. After the alignment, the CT and CBCT images were fitted perfectly, therefore, the initial spatial variation between the CBCT and CT images was considered to be zero. After transforming according to the affine matrix, the CBCT images were used as the floating images. By employing the same method, 15 pairs of standard images were derived. The standard images were analyzed using different algorithms (IBR, CNNR and CIR). Those algorithms were evaluated by time complexity and registration accuracy. The time complexity of three algorithms was calculated by the average running time. The registration accuracy was measured by the similarity between the floating images and the fixed images after alignment. The TREs and set-up error were used to evaluate the similarity of the inputted image pairs. Each image pair was calculated for 100 times. TREs were calculated by the residual errors of the results. The running time is the average value of 100 computations. The set-up error are the statistical results of all the calculations.

The running time of different algorithms are shown in the Fig. 7. The CNNR was calculated using GPUs (Nvidia Quadro P5000) and the IBR was calculated using multi-core CPUs (Intel Xeon E5-2650 with 2.20 GHz and 128 GB RAM). The CIR was calculated by both the CPUs and GPUs. The average running time for CNNR is 0.21 s, for CIR is 8.3 s, and for IBR it took remarkable longer time of 62.7 s.

The accuracy, which is measured by TREs are shown in fig. 8. The average TREs of CNNR, IBR and CIR are 3.3, 5.2 and 0.51, respectively. The standard deviations are ±0.15, ±1.0, ±0.08, respectively. The results suggest that CIR has greatly improved the accuracy.

Moreover, the average set-up errors in translation and rotation around the X, Y, and Z axes are shown in fig. 9. The set-up errors of CIR are within 0.5 mm and 0.2 °. For CNNR the set-up errors are within 3.0 mm and 0.8 °, and for IBR are within 4.5 mm and 1.8 °. CIR has the lowest set-up error, that confirms its highest accuracy.

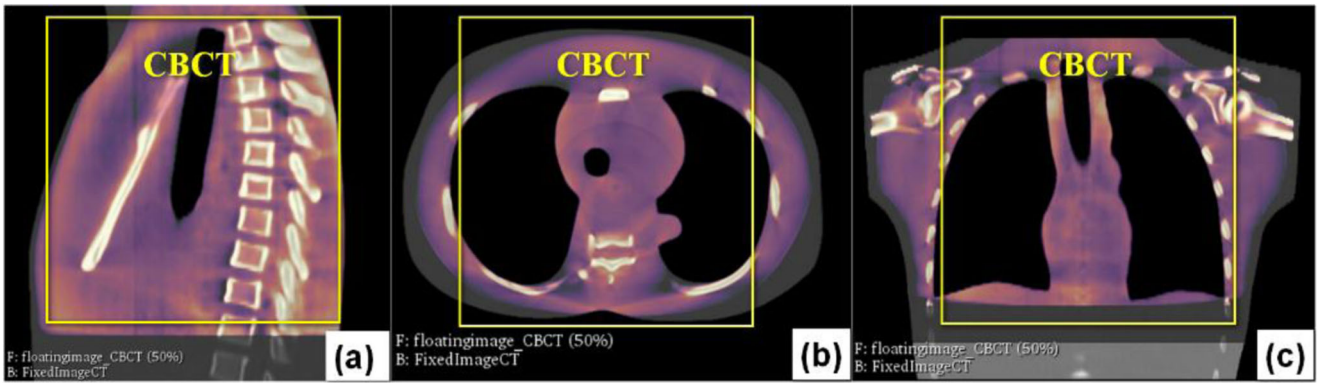

Fig. 6 Images taken from different directions for a chest phantom: a) Transverse; b) Sagittal; c) Coronal. The CBCT images are in color (in yellow square) with 50% transparency, and the CT images are in gray

Discussion

The goal of radiation treatment is to deposit sufficient dose to kill the cancer cells while affect as little as possible the surrounding healthy tissues. In practice, due to the anatomical complexity, especially when a tumor locates near neurological structures such as the brain stem or spinal cord, it is of crucial importance to precisely identify the target region, for which the alignment of planning CT and CBCT plays a key role. For the intensity-based registration, its cost function which is associated with intensity similarity metrics is often non-convex [18]. If the capture range is wide and the input images are complex, the intensity-based registration is not able to give an accurate result. On the other hand, the CNNs can find the approximate pose of the input 3D volume and overcome the non-convex issues. However, the CNNs has its own bottlenecks due to a lack of ground truth.

Our new algorithm is a hierarchical registration strategy, it takes advantages of the CNNs and upgrades the conventional registration algorithm. The CNNS severs as a coarse registration. Spatial variation between the floating and fixed images is calculated by neural network model without using iterative optimization. The rough result calculated by CNNs smooths the spatial variation. The search range of the intensity-based registration is reduced, while the performance of the intensity similarity metrics is improved. For the intensity-based method, it is used as fine registration to eliminate the residual set-up error. Learning-based registration is employed in order to promote the performance of the intensity-based registration. The accuracy of the CNNS is promoted, too.

One of the main concerns with this new method is the accuracy of the CNNR. The CNNR gives out a rough result in a short time, but its accuracy is limited due to the size of training datasets (24,000 volume in this work). With a better training to the neural network model, the accuracy of CNNR should be further improved, and the running time of the CIR will be significantly shortened.

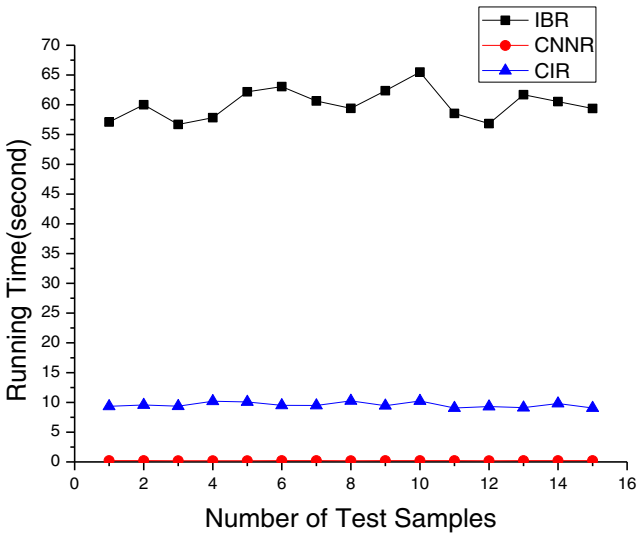

Fig. 7 The running time of the IBR, CNNR and CIR

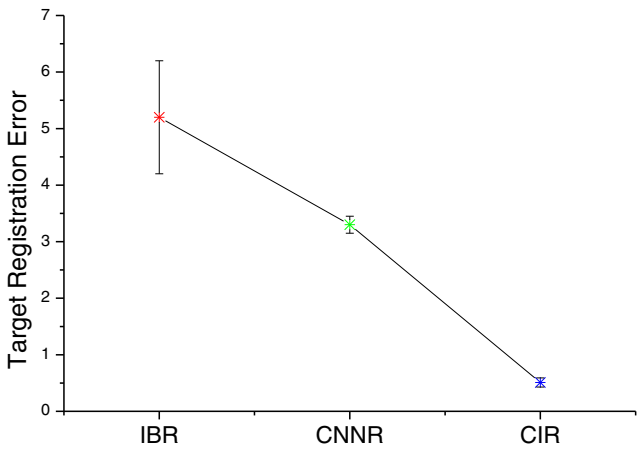

Fig. 8 The calculated TREs of the IBR, CNNR and CIR

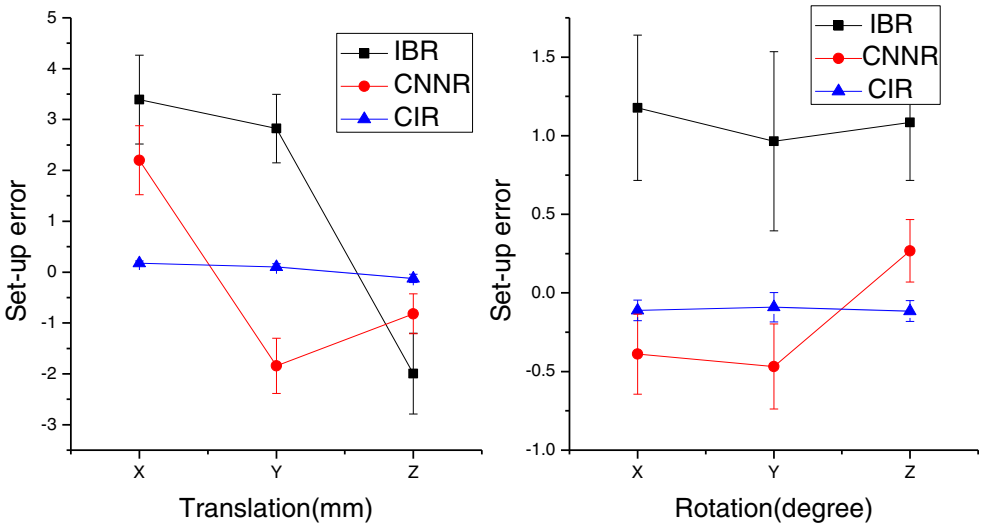

Fig. 9 The set-up errors of rotation and translation for the IBR, CNNR and CIR

Conclusion

In this paper, a new method which combines the conventional algorithm with the convolution neural networks is proposed for IGRT. Within the capture range of the treatment couch, this new method shows its advantages compares to the state-of-art image registration algorithm. The set-up error is reduced to sub-millimeter level.

Pairs of CBCT and CT images are employed as the basic training data to train the neural network model. After processing by the oncologist, the image pairs provide the ground truth that overcome a major drawback in using the CNNs for image registration. Notice that the image pairs are taken from four different parts of the cancer patients, it guarantees the generalization of neural network model during IGRT. TRE is used as the loss function during regression, and it improves the computational accuracy of every iteration of back-propagation.

The combination of the convention algorithm and convolution neural networks take advantage of both, there is still some space for further reducing the running time, the framework of the network should be fine-tuned. The promoted CNNs can give out an accurate result rapidly, the intensity-based registration works as the last safeguard, so our new method can align the planning CT images with the CBCT images accurately in real time during IGRT.

Acknowledgments We would like to thank Cancer Hospital, Chinese Academy of Science for providing CT and CBCT image datasets. This work was supported by Key Program of 13th five-year plan, CASHIPS (Grant No. KP-2017-24).

Compliance with ethical standards

Conflict of Interests The authors have no conflict of interests regarding the publication of this paper.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed Consent For this type of study formal consent is not required.

References

1. Sorcini, B., and Tilikidis, A., Clinical application of image-guided radiotherapy, IGRT (on the Varian OBI platform). Cancer/Radiothérapie 10(5):252–257, 2006.
2. Sharma, S. D., Dongre, P., Mhatre, V., and Heigrujam, M., Evaluation of automated image registration algorithm for image-guided radiotherapy (IGRT). Australas Phys Eng S 35(3):311–319, 2012.
3. Boda-Heggemann, J., Lohr, F., Wenz, F., Flentje, M., and Guckenberger, M., kV cone-beam CT-based IGRT. Strahlenther Onkol 187(5):284–291, 2011.
4. Johansen, J., Bertelsen, A., Hansen, C. R., Westberg, J., Hansen, O., and Brink, C., Set-up errors in patients undergoing image guided radiation treatment. Relationship to body mass index and weight loss. Acta Oncol 47(7):1454–1458, 2008.
5. Guckenberger, M., Meyer, J., Wilbert, J., Baier, K., Sauer, O., and Flentje, M., Precision of image-guided radiotherapy (IGRT) in six degrees of freedom and limitations in clinical practice. Strahlenther Onkol 183(6):307–313, 2007.
6. McLaughlin, R. A, Hipwell, J., Penney, G. P., Rhode, K., Chung, A., Noble, J.A., Hawkes, D., Intensity-based registration versus feature-based registration for neurointerventions. Proc Med Image Understanding Analysis (2001)
7. Matsopoulos, G. K., Mouravliansky, N. A., Delibasis, K. K., and Nikita, K. S., Automatic retinal image registration scheme using global optimization techniques. Ieee T Inf Technol B 3(1):47–60, 1999.
8. Rouet, J.-M., Jacq, J.-J., and Roux, C., Genetic algorithms for a robust 3-D MR-CT registration. Ieee T Inf Technol B 4(2):126–136, 2000.
9. Brounstein, A., Hacihaliloglu, I., Guy, P., Hodgson, A., Abugharbieh, R., Towards real-time 3D US to CT bone image registration using phase and curvature feature based GMM matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2011. Springer, pp 235-242

10. Greenspan, H., Van Ginneken, B., and Summers, R. M., Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. Ieee T Med Imaging 35(5): 1153–1159, 2016.
11. Krizhevsky, A., Sutskever, I., Hinton, G. E., Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, 2012. pp 1097-1105
12. Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., Comaniciu, D., An Artificial Agent for Robust Image Registration. In: AAAI, 2017. pp 4168-4175
13. Salehi, SSM., Khan, S., Erdogmus, D., Gholipour, A., Real-time Deep Registration With Geodesic Loss. arXiv preprint arXiv: 180305982, (2018)
14. Mahendran S, Ali H, Vidal R 3D pose regression using convolutional neural networks. In: IEEE International Conference on Computer Vision, 2017. vol 2. p 4
15. Chou, C. R., Frederick, B., Mageras, G., Chang, S., and Pizer, S., 2D/3D image registration using regression learning. Comput Vis Image Und 117(9):1095–1106, 2013. https://doi.org/10.1016/j.cviu.2013.02.009.
16. Khamene, A., Bloch, P., Wein, W., Svatos, M., and Sauer, F., Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. Med Image Anal 10(1):96–112, 2006.
17. Barber, J., Sykes, J. R., Holloway, L., and Thwaites, D. I., Comparison of automatic image registration uncertainty for three IGRT systems using a male pelvis phantom. J Appl Clin Med Phys 17(5):283–292, 2016.
18. Sloan, J., Goatman, K., Siebert, J. Learning rigid image registration-utilizing convolutional neural networks for medical image registration (2018)