Mathematical Biosciences and Engineering

MBE, 20(3): 4403–4420.
DOI: 10.3934/mbe.2023204
Received: 10 October 2022
Revised: 16 November 2022
Accepted: 20 November 2022
Published: 23 December 2022

http://www.aimspress.com/journal/MBE

Research article

Enhancement of cone beam CT image registration by super-resolution pre-processing algorithm

Liwei Deng[1], Yuanzhi Zhang[1], Jingjing Qi[1], Sijuan Huang[2], Xin Yang[2,*] and Jing Wang[3,*]

[1] Heilongjiang Provincial Key Laboratory of Complex Intelligent System and Integration, School of Automation, Harbin University of Science and Technology, Harbin 150080, China.
[2] Department of Radiation Oncology; Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, China.
[3] Faculty of Rehabilitation Medicine, Biofeedback Laboratory, Guangzhou Xinhua University, Guangzhou 510520, China.

* **Correspondence:** Email: happyjing00@163.com, yangxin@sysucc.org.cn.

Abstract: In order to enhance cone-beam computed tomography (CBCT) image information and improve the registration accuracy for image-guided radiation therapy, we propose a super-resolution (SR) image enhancement method. This method uses super-resolution techniques to pre-process the CBCT prior to registration. Three rigid registration methods (rigid transformation, affine transformation, and similarity transformation) and a deep learning deformed registration (DLDR) method with and without SR were compared. The five evaluation indices, the mean squared error (MSE), mutual information, Pearson correlation coefficient (PCC), structural similarity index (SSIM), and PCC + SSIM, were used to validate the results of registration with SR. Moreover, the proposed method SR-DLDR was also compared with the VoxelMorph (VM) method. In rigid registration with SR, the registration accuracy improved by up to 6% in the PCC metric. In DLDR with SR, the registration accuracy was improved by up to 5% in PCC + SSIM. When taking the MSE as the loss function, the accuracy of SR-DLDR is equivalent to that of the VM method. In addition, when taking the SSIM as the loss function, the registration accuracy of SR-DLDR is 6% higher than that of VM. SR is a feasible method to be used in medical image registration for planning CT (pCT) and CBCT. The experimental results show that the SR algorithm can improve the accuracy and efficiency of CBCT image alignment regardless of which alignment algorithm is used.

Keywords: super-resolution; cone-beam CT; image registration; deep learning; medical image processing

1. Introduction

Medical image registration is an important branch of computer vision [1,2]. It is a significant step in actualizing medical image analysis, completing auxiliary clinical diagnosis, and understanding medical images [3–5]. Registration is also the basis of fusion [6]. Medical image registration is a necessary step in image guidance, motion tracking, and image segmentation [7]. It is also widely used in the comparison of real medical images and atlases, surgical navigation, tumor parameter estimations, cardiac motion estimations, the creation of an average atlas, surgical location, and radiotherapy plan design [8–10].

In modern external beam image-guided radiation therapy (IGRT) [11], cone-beam computed tomography (CBCT) is commonly used to monitor daily anatomical changes [12]. The image quality of CBCT is inferior to that of computed tomography (CT) for soft tissues due to the presence of artifacts. Therefore, deformable registration (DR) of planning CT (pCT) to the daily anatomy of CBCT images has been proposed to correct CBCT imaging artifacts for adaptive radiation therapy [13–15]. CBCT can monitor the location of a tumor in real-time [16,17]. With the help of CBCT, the radiation dose of normal tissue can be reduced, and the radiation dose of the tumor area can be increased, which can improve the local control rate of the tumor and reduce radiotherapy complications [18].

How to improve the quality of CBCT images for registration is an important problem [19]. A number of scholars have proposed automatic registration methods, which have greatly improved the efficiency and accuracy of medical diagnosis and treatment [20]. However, image quality seriously affects readability because CBCT is greatly affected by scattered rays contains considerable noise, artifacts and blurry edges [21,22]. There remains a large difference between the low contrast areas of CBCT images and conventional pCT images. It is especially difficult to recognize and outline target areas.

Deep learning is involved in considerable image registration research [23,24]. Broadly speaking, two strategies are prevalent in the literature: 1) Wu et al. [25], Simonovsky et al. [26], and Cheng [27] used deep learning networks to estimate a similarity measure for two images to derive an iterative optimization strategy; and 2) Miao et al. [28] and Yang et al. [29] directly predicted transformation parameters using deep regression networks. Recently, DeTone [30] proposed a homography network, which is a supervised network similar to the visual geometry group [31]. The algorithm learns the advantages of the homography and the convolutional neural network model parameters in an end-to-end way. Krebs [32] used artificial agents to optimize the parameters of the deformation models, and good results were obtained in 2D/3D. Currently, compared to medical images with accurate annotations, medical image data without annotations are easier to obtain, and unsupervised learning has quickly become a hot research topic [33]. In an unsupervised registration network, registration pairs are imported into the network to obtain a deformation field, and the moving image is transformed by deformation interpolation to obtain the registered image. In an unsupervised registration network, the selection and usage of the similarity measure function is an important and difficult point to improve the registration accuracy.

In our proposed deformed image registration network, a total of five similarity functions are used,

namely, the mean squared error (MSE), mutual information (MI), Pearson correlation coefficient (PCC), structural similarity index measurement (SSIM) and PCC + SSIM.

Here, we propose a deep learning deformed registration (DLDR) method using super-resolution preprocessing (SR-DLDR). The method uses a high precision super resolution (SR) approach based on a very deep super resolution (VDSR) network. Next, three traditional rigid registration methods (rigid, affine and similarity) prove that adding SR during image preprocessing can significantly improve the accuracy of rigid body registration. Finally, experimental comparisons between registration with and without SR are given. Furthermore, we also improved the similarity function for the DR by combining the PCC with SSIM as the similarity measure function of the model. The experimental results show that SR-DLDR is superior and more reliable than VoxelMorph (VM).

In short, using the nature of SR [34–36] image processing in the denoising stage of CBCT and using the VDSR model can improve the image quality and fill in the blank information of an image. The quantitative and qualitative indicators of CBCT images and pCT image registration after preprocessing are better.

2. Methods

2.1. SR in CBCT image preprocessing

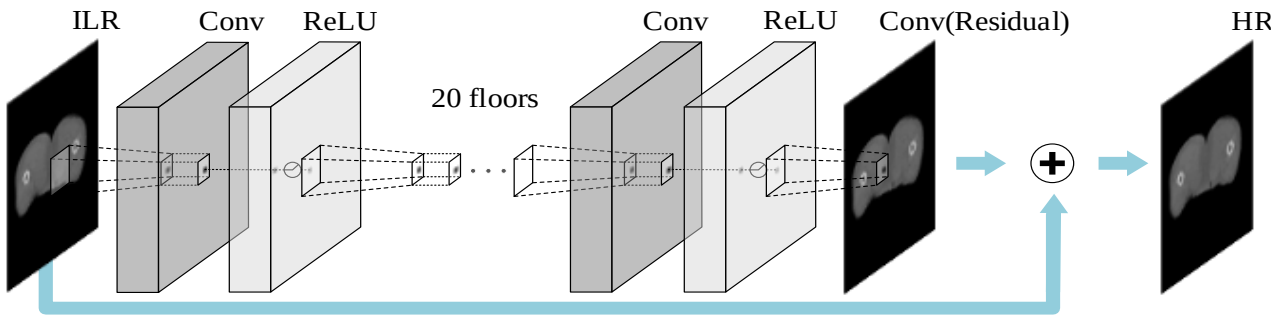

Figure 1. Proposed SR network structure. The network repeatedly cascades a pair of convolutional and nonlinear layers. An interpolated low-resolution (ILR) image goes through layers and is transformed into a high-resolution (HR) image. The desired output (HR) image is the addition of the input (ILR) image and the residual image predicted by the SR network.

Nearly 400 sectioned images were used in this study. The image dataset of each patient was obtained before the radiotherapy fraction. Each patient's dataset contained a group of 120 slices of pCT images and one group of 90 slices of CBCT images. The pCT images had a matrix size of 512 by 512 on the axial plane with a pixel size of 1.171875 mm by 1.171875 mm. The CBCT images had a matrix size of 410 by 410 on the axial plane with a pixel size of 1 mm by 1 mm. The thicknesses of the CT and CBCT slices were both 3 mm. Among the images from the five patients, those from four patients were used to train the network, and those from the remaining patients were retained for testing. Using the leave-one-out cross validation method, four patients were randomly selected from the five patients datasets as the training set, and the rest were used as the test set. This process was repeated

five times. The final result was the mean value of five verifications. To accelerate the model training time, we used paired data for modeling. We also clipped all image sizes down to 384 × 384 to minimize the anatomical region to accelerate the calculation time.

Figure 1 shows the SR model flow of CBCT enhancement processing [37]. It is a single image super-resolution processing model with a very high recovery rate. The model uses 20 weight layers to improve the accuracy and the network architecture is deepened significantly. And then by cascading small filters several times, the contextual information of large image regions is applied more comprehensively in the deep learning network. However, for such a deep network model, the training time is significantly longer and the training cost increases. In response, we propose a simple and effective training procedure to control the training time.

The SR system model was constructed by using a forward relation model [37]. The relationship between a low-resolution image and a high-resolution image is expressed as follows:

$$y_k = DB_j M_{k,j} x_j + n_j, 1 \le j, k \le p \tag{1}$$

where p is the number of frames in the image sequence; x_j, y_k and n_j are the high-resolution image of the j-th frame to be calculated, the low-resolution image observed in the k-th frame and the noise during image acquisition, respectively; D is the down-sampling matrix; B_j is the blur matrix; and $M_{k,j}$ is the motion matrix composed of the motion vectors between the j-th and k-th frames.

This network was inspired by Simonyan and Zisserman [38]. Twenty weighted layers were utilized. The model has three layers: reconstruction, nonlinear mapping, and patch extraction/representation. The SR network predicts picture details from an interpolated low-resolution image as input. In SR approaches the technique modeling image details is frequently used. The SR model consists of 20 layers, all of which, with the exception of the first and last, are of the same type: a $3 \times 3 \times 64$ filter that acts on the 3×3 spatial region using 64 channels. The input image is used by the first layer. The last layer is a single $3 \times 3 \times 64$ filter that is used to recreate images. At the input of the model, the image size is processed by MATLAB. MATLAB processes the image size at the model's input. We uniformly reduced the low-resolution photos from an image block size of 384×384 down to 123×123 using bicubic interpolation, and then we fed the low-resolution images into the model to train and output the predicted image detail texture [39]. To draw conclusions using the SR approach, image center pixels need a large number of surrounding pixels. A broad surrounding area will result in poor inference accuracy for the center pixels; therefore, it should be kept to a minimum. A final image that has been cropped is too small to be aesthetically attractive. We make use of contextual data dispersed across very vast image regions. It frequently happens that a little patch of information is insufficient to recover details (ill-posed). A VDSR network with a wide receptive field takes into account a lot of the visual context.

2.2. Deformable registration

The moving image I_m and the fixed image I_f in the model frame diagram are and respectively. To create a predictive registration domain using deep learning, the model is fed a pair of moving and stationary photos. To create the distorted moving image, the moving image is subjected to a registration field and B-spline space transformation. Calculating the difference between the fixed image and the distorted images allowed us to return the loss difference to the deformed image registration network,

which we then used to retrain the deformed field until the loss difference was minimum and the network was at its best [40]. The loss function can be expressed mathematically as follows:

$$\begin{aligned} l(I_m, I_f, \phi; \theta) &= l_{sim}(I_m \circ \phi, I_f; \theta) + \lambda R(\phi; \theta) \\ &= l_{sim}(I_m \circ f_\theta(I_m, I_f), I_f; \theta) + \\ &\quad \lambda R(f_\theta(I_m, I_f), \theta) \end{aligned} \tag{2}$$

where l_{sim} is the image similarity measure and R denotes the regularization of ϕ. λ is a controllable weighting parameter that the user defines. The minimizer can then estimate the parameters θ that produce the ideal registration field [41].

$$\theta^+ = \arg\min_\theta l(I_m, I_f, \phi; \theta) \tag{3}$$

The optimal ϕ is then given by:

$$\phi^+ = f_\theta * (I_m, I_f) \tag{4}$$

2.3. SR-DLDR

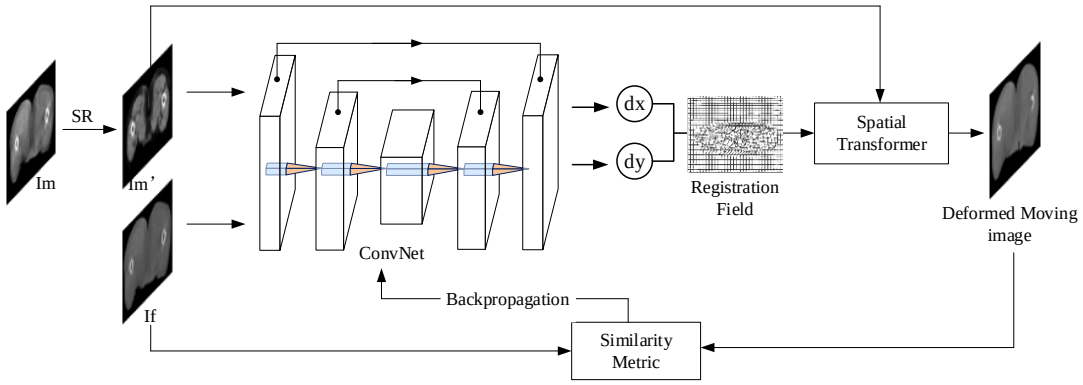

Figure 2. Registration process of CBCT and pCT. The CBCT image is used as the moving image I_m, and $I_m^{'}$ is obtained after SR processing; then, the image pair is synthesized with the pCT image I_f and input into the alignment network. The B-spline spatial transformer and registered field are then used to transform the moving image CBCT. The ConvNet's parameters are updated by computing the similarity measure loss between the deformed image and the fixed image pCT.

The SR-DLDR process is shown in Figure 2. First, the CBCT image is processed using the SR processing algorithm described in the previous section, and the resulting high-quality CBCT image is used as moving image I_m. After that, the pCT image is cropped to match the CBCT size and used as the fixed image I_f. The moving image I_m and fixed image I_f are combined as a set of image pairs to be the input of the alignment network, and a spatial deformation field is generated by a U-shaped convolutional network to complete deformation of the moving image I_m. A spatial transformer is used

to distort the moving image I_m according to the generated deformation field data to obtain the alignment image. The similarity measure between the aligned image and the fixed image is calculated as a loss and returned to the U-shaped convolutional network to update the network parameters to obtain the best aligned deformation field through continuous iteration.

2.3.1. ConvNet structure

The network structure of ConvNet is similar to that of U-Net [42]. The input size of the model is 384×384, and the output image size is the same. The network has a coding path to transform a single input image pair into a $2 \times M \times M$ volume. The convolutional layer of each layer is followed by a rectified linear unit (ReLU), and the 2×2 maximum pooled downsampling method is adopted. In the decoding stage, the upsampling "up-convolution" calculation is used [42]. The output structure of this part is integrated with the output of the feature graph in the coding process by upsampling each feature graph. The output registration domain with 16 feature maps is then created by using 16 3×3 convolutions and two 1×1 convolutions. Figure 2 depicts the network architecture schematically.

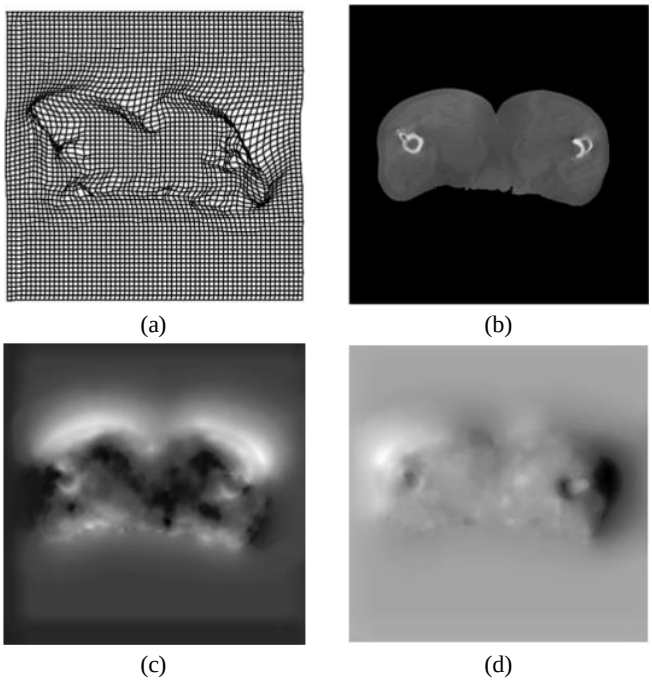

Figure 3. Images showing (a) the deformed grid map of the registration model training output, (b) the deformed image of the moving image, (c) the visualization effect of the input image of the 11th layer, and (d) the input image processed by the convolution layer in the network training process.

2.3.2. B-spline spatial transformation

Nonlinear picture deformation is applied by using the B-spline function as the spatial transformer [41,43]. By combining the effects of the B-spline spatial converter and the deformation registration domain, the moving picture is used to calculate the loss error with the fixed reference

image [44], as shown in Figure 3. The definition of the B-spline basis function is as follows:

$$N_{i,0}(u) = \begin{cases} 1 & if\ u_i \le u < u_{i+1} \\ 0 & otherwise \end{cases}$$
$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \tag{5}$$

Set u as a collection of $m+1$ nonrecurring subtractions, $u_0 \le u_2 \le u_3 \le u_4 \le ... \le u_m$, u_i as the nodes, u as the node vector, and the semi-open interval $[u_i, u_{i+1})$ as node interval i. Note that some u_i may be equal and some node intervals will not exist. If a node u_i appears k times (i.e., $u_i = u_{i+1} = ... = u_{i+k-1}$), where $k > 1$, u_i is a multiple node with a repeatability of k, and it is written as $u_i(k)$. p represents the number of base functions.

The warp is determined by a flow field of displacement vectors ω that define the correspondences of pixel intensities in the output image to the pixel locations in the moving image. The intensity at each pixel location M in the output image $I_m \circ \phi(M)$ is defined by:

$$I_m \circ \phi(M) = I_m(M - \omega(M)) \tag{6}$$

2.4. Similarity measure function

The similarity measure, which is used to evaluate the outcomes of each transformation and serve as the foundation for the following step in the search strategy, is a criterion for measuring the results of each transformation.

The feature space and search space are strongly related to similarity transformation, and various feature spaces frequently correlate to various similarity measures. The value of the similarity measure [45] will be used to choose the registration transformation and decide whether the image is appropriately matched with the present transformation model. The capacity of the registration method to resist jamming is often assessed using feature extraction and similarity measurement.

In our work, we applied five evaluation indices to verify the effect of image registration. They are the MSE, MI, PCC, SSIM and PCC+SSIM.

2.4.1. MSE

The MSE is the sum of the average values divided by the square of the difference between the actual value and the anticipated value. The square method is frequently employed as the loss function in linear regression since the calculation is straightforward to lead. A measurement of image fidelity that works for both fixed and moving images with a similar contrast and intensity distribution is the MSE. The MSE is calculated as

$$MSE(I_d, I_f) = \frac{1}{\Omega} \sum_{i \in \Omega} \| I_f(i) - I_d(i) \|^2 \tag{7}$$

where Ω is the spatial domain I_d and I_f represent the alignment image and the target image, respectively. In order to ensure the consistency of the loss function in the model output, the similarity function is defined as: $l_{sim}(I_m, I_f, \phi; \theta) = MSE(I_d, I_f)$.

2.4.2. MI

In information theory, MI is a valuable information metric [46]. It can be thought of as the amount of knowledge that one random variable has about another, or as the uncertainty of one random variable being lessened by being aware of another. It is frequently important to compare the similarity of two photos when processing images. Consider two photos, A and B. Then, the calculation formula of their mutual information value is:

$$I(A, B) = H(A) + H(B) - H(A, B) \tag{8}$$

$H(A, B)$ is the joint entropy of A and B. Let $pI_f(a)$ and $pI_d(b)$ be the marginal probability distributions of the fixed and deformed moving images. MI can be defined as follows:

$$MI(I_d, I_f) = \sum_{a,b} pI_f I_d(a, b) \log \frac{pI_f I_d(a, b)}{pI_f(a).pI_d(b)} \tag{9}$$

The joint distribution, $pI_f I_d(a, b)$, can be computed as follows

$$pI_f I_d(a, b) = \frac{1}{\Omega} \sum_{i \in \Omega} \delta(I_f(i) - a) \delta(I_d(i) - b) \tag{10}$$

The loss function produced using MI as a similarity measure function cannot be back-propagated to the network because the Dirac delta function is not differentiable; thus, we use the differentiable Gaussian function to roughly substitute $pI_f I_d(a, b)$.

$$pI_f I_d(a, b) = \frac{1}{\Omega} \sum_{i \in \Omega} \frac{1}{\sigma^2 2\pi} e^{\frac{(I_f(i) - a)^2}{\sigma^2 2}} e^{\frac{(I_d(i) - b)^2}{\sigma^2 2}} \tag{11}$$

where σ is a user-defined parameter that can vary depending on the images of a certain application. The calculated MI value is negative, so the loss function is defined as: $l_{sim}(I_m, I_f, \phi; \theta) = -MI(I_d, I_f)$.

2.4.3. PCC

The Pearson product-moment correlation coefficient, which has a value of $[-1, 1]$, is a statistical measure of the correlation (or linear correlation) between two variables x and y. The coefficient is frequently used in natural science to assess how closely two variables are correlated with one another. Its use in the registration of medical images has been discussed [47]. The PCC is defined as the

covariance between images divided by the product of their standard deviations:

$$PCC(I_d, I_f) = \frac{\sum_{i\in\Omega}(I_f(i)-\bar{I}_f)(I_d(i)-\bar{I}_d)}{\sqrt{\sum_{i\in\Omega}(I_f(i)-\bar{I}_f)}\sqrt{\sum_{i\in\Omega}(I_d(i)-\bar{I}_d)}} \tag{12}$$

where $\bar{I}_f$ and $\bar{I}_d$ represent the mean intensities, Ω is the total number of pixels in the image, and $I_*(i)$ is the pixel value at the corresponding position in the image. In order to lose the consistency of the function, let $l_{sim}(I_m, I_f, \phi; \theta) = 1 - PCC(I_d, I_f)$.

2.4.4. SSIM

Brightness, contrast, and structure are the three main characteristics that the structural similarity index may derive from an image [48]. From the standpoint of picture composition, distortion is characterized as a mix of brightness, contrast, and structure, whereas structural information is defined as a property independent of the brightness and contrast to reflect the structure of the objects in a scene. The standard deviation is used to evaluate contrast, the covariance is used to quantify structural similarity, and the mean is used to estimate brightness. The SSIM is a crucial metric for assessing the quality of medical image registration.

$$SSIM(I_d, I_f) = \frac{(2\mu_{I_d}\mu_{I_f} + C_1)(2\sigma_{I_f I_d} + C_2)}{(\mu_{I_f}^2 + \mu_{I_d}^2 + C_1)(\sigma_{I_f}^2 + \sigma_{I_d}^2 + C_2)} \tag{13}$$

where C_1 and C_2 are small constants needed to avoid instability; μ_{I_f} and μ_{I_d}, σ_{I_f} and σ_{I_d} are the local standard deviations of images I_f and I_d, respectively. Similarly, in order to ensure the consistency of the loss function, the loss function is defined as: $l_{sim}(I_m, I_f, \phi; \theta) = 1 - SSIM(I_d, I_f)$.

2.4.5. PCC + SSIM

The PCC is more sensitive to robust image noise but less sensitive to fuzzy edges. Using the PCC as the loss function alone causes slow convergence, and the SSIM can model image details, including noise and artifacts, through the network. Therefore, we combine the PCC and SSIM to evaluate the image registration performance. Both the PCC and SSIM [40] are bounded within the range $[-1,1]$. The closer the value is to 1, the higher the similarity between the two images. Thus, we propose combining the SSIM and PCC using equal weights:

$$\begin{aligned} l_{sim}(I_m, I_f, \phi; \theta) = {} & 0.5 * (1 - SSIM(I_d, I_f)) \\ & + 0.5 * (1 - PCC(I_d, I_f)) \end{aligned} \tag{14}$$

3. Experiments and results

The goal of the work was to improve CBCT with pCT registration accuracy. We performed validation on pelvic data from clinical patients and selected a total of 400 pairs of section data from the corresponding sites. Of these, 350 slice pairs were selected as the training set and 50 slice pairs as

the test set. The image size was based on the SR-processed CBCT, and all slices were processed to a 384 × 384 size. The range of selected pixel values was between [0–2000 HU]. The proposed method was implemented by using Keras with a TensorFlow backend on an NVIDIA 2080TI (11 GB). The platform environment of the experiment was unchanged, and the dataset used was that of the image after the SR process was performed. The registration results were compared with those of VoxelMorph (VM).

For the parameter setting of the network, due to the hardware device limitation of the training platform, we set the learning rate to 10^{-4} and the learning rate decay to 0.88. The input images were not sliced, the number of single batch inputs was set to 8 and the number of iterations was set at 4000. After extensive testing, we were able to basically determine that the alignment network will reach convergence at about 3920 iterations in the worst case.

3.1. SR + Rigid registration validation and evaluation

First, we looked at how SR pretreatment affected the stiff alignment of medical images. The studies showed that SR-treated samples performed better under rigid alignment. Figure 4 compares the three registration maps with CBCT preprocessing to the registration maps with affine transformation, comparable transformation, and stiff transformation. Table 1 lists five distinct quantitative evaluation indicators of the similarity measure function in addition to the qualitative analysis of the benefits and drawbacks of registration in Figure 4.

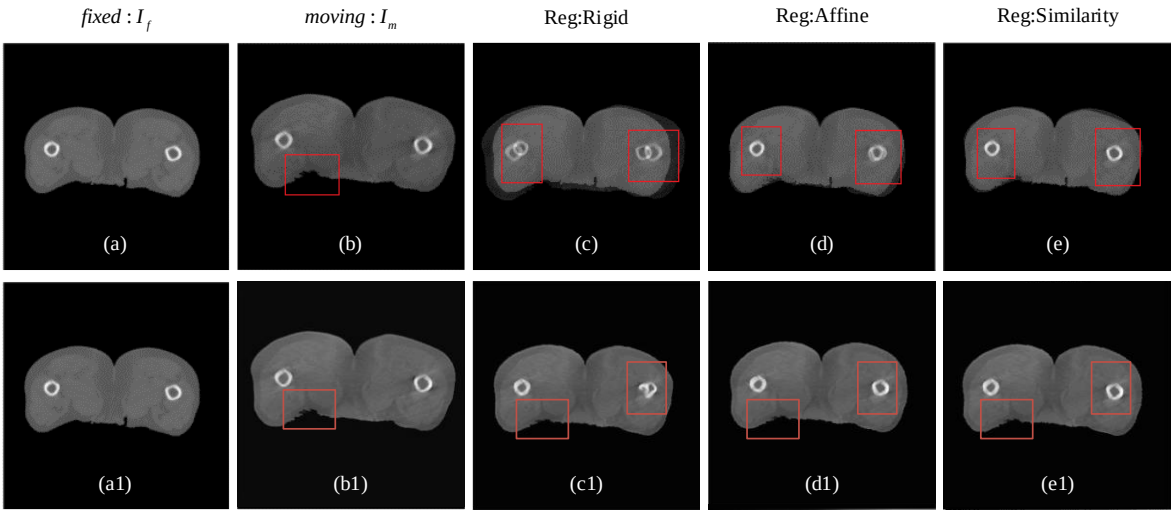

Figure 4. Comparison of the alignment effect of the original sample and the SR sample. The rigid transform, affine transform, and related transform pictures of the original sample are displayed in the first row. After SR processing, the sample's rigid transform, affine transform, and comparable transform images are registered in the second row. (a) and (a1) are fixed images, and (b) and (b1) are moving images. (c) and (c1) are the rigid transformed results, (d) and (d1) are the affine transformed results, and (e) and (e1) are the similar transformed results.

Through observation using human vision, the experimental results in this paper vaguely show that the effect of rigid registration of medical images is not good; therefore, it was gradually replaced by

elastic registration. In the experiment, four image registration similarity indices (the MSE, MI, PCC and SSIM) were used to compare the rigid registration effect. "SR + Rigid" means rigid transformation registration based on SR, "SR + Affine" means affine transformation registration based on SR, and "SR + Similarity" means similarity transformation registration based on SR. The data in Table 1 show that the second method has higher accuracy in medical image registration.

As shown in the picture effect, the first line of Figure 4 starts from the third column, which is an affine transformation registration map, rigid transformation registration map, and similar transformation registration map, respectively. Combined with quantitative data comparison, it can be concluded that the similarity transformation of rigid registration is better than affine transformation and better than rigid transformation. The results of the comparison of experimental data with and without preprocessing show that the accuracy of adding the preprocessing method is higher than that of direct registration.

Table 1. Quantitative rigid registration evaluation index with and without SR.

Method/Metric	MSE	MI	PCC	SSIM
Rigid transformation	0.6983	0.4857	0.7230	0.7245
SR + Rigid transformation	0.6542	0.5245	0.7880	0.7644
Affine transformation	0.5666	0.5683	0.8275	0.7980
SR + Affine transformation	0.5427	0.5914	0.8584	0.8137
Similarity transformation	0.5539	0.5940	0.8574	0.8048
SR + Similarity transformation	0.5211	0.6014	0.8765	0.8128

3.2. SR + DR validation and evaluation

3.2.1. DR comparison with and without SR

The registration examples of five different similarity measure loss functions are shown in Figure 5. The top row is the direct image registration results through the ConvNet under five loss functions, and the second row is the resulting images of CBCT registration with pCT after SR preprocessing. In addition, Table 2 shows the results.

In the experiment, "an iteration" represents the number of times the network is updated by back-propagation. It is also the number of registered domain updates. According to the experiments performed in this paper, the effect of the model is the best when the number of iterations is 4000.

In Table 2, "Iterations" denotes the number of times the parameters are updated by the registered network model. There are five types of loss functions. Table 2 lists the registration effects under different numbers of iterations (100, 500, 1000 and 4000) and different evaluation functions in detail. At 4000 iterations, the registration effect of SR-DLDR is better than that of DLDR under each evaluation index. In Table 2, bold fonts are used to express excellent effects.

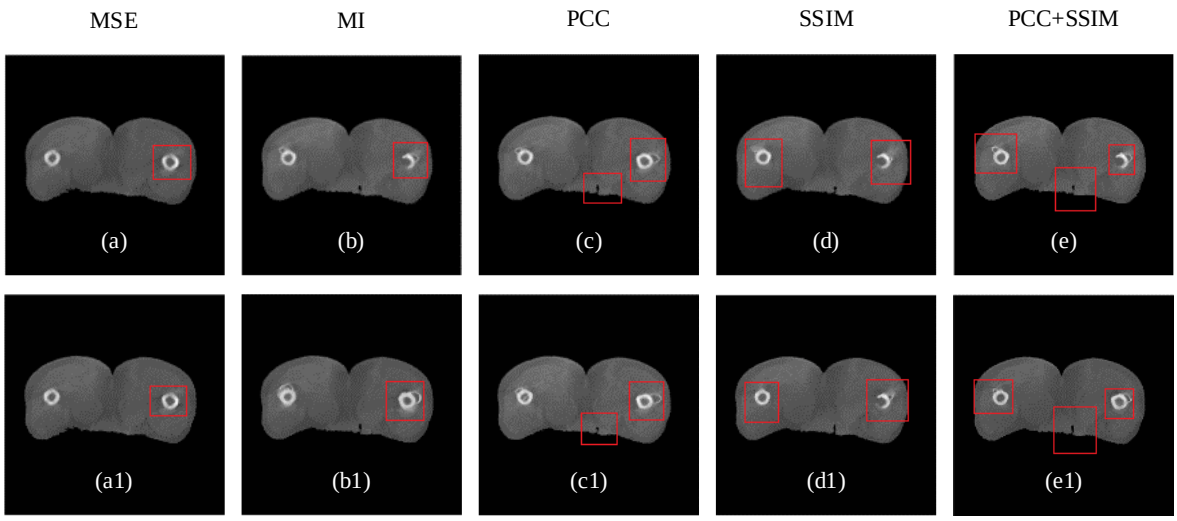

Figure 5. The first row of images shows the registration of CBCT and pCT without SR processing. From (a) to (e), there are five different similarity measure functions: the MSE, MI, PCC, SSIM and PCC + SSIM, respectively. The second row from (a1) to (e1) corresponds to each of the above loss functions, which are the registration results of CBCT and pCT with SR processing.

Table 2. Quantitative DR evaluation indices with and without SR for five loss functions and different numbers of iterations.

Loss functions	Iterations			
	100	500	1000	4000
MSE	0.0124225	0.0069414	0.0041089	0.0022588
SR + MSE	0.0121268	0.0072499	0.0039648	0.0022189
MI	0.5468396	0.5826355	0.5879228	0.5957638
SR + MI	0.5412521	0.5856629	0.5879228	0.5979443
PCC	0.9692372	0.9819638	0.9875129	0.9900116
SR + PCC	0.9643641	0.9817635	0.9884341	0.9900189
SSIM	0.9011286	0.9474509	0.9557047	0.9626334
SR + SSIM	0.9010728	0.9534056	0.9625438	0.9666951
PCC + SSIM	0.9412062	0.9658692	0.9658692	0.9727479
SR + PCC + SSIM	0.9338375	0.9638376	0.9699244	0.9770569

In each iteration, by comparing the deformation of the moving image with the loss function of the fixed image, the corresponding difference will be obtained. In the experiment, the difference of each iteration was retained. To output the experimental results with high clarity and readability, the average loss functions of 500 iterations, 1000 iterations and 4000 iterations were taken as the experimental results. The advantage of the data is not obvious when the MSE and MI are used as evaluation indices. Only when the number of iterations was more than 1000 was the loss function of

SR-DLDR slightly better numerically than common registration without SR processing. When the SSIM and PCC combination was used as a loss function, the image registration accuracy was better. As the evaluation index of structural similarity, the output result of SR-DLDR was better than that of common registration without SR processing from the beginning, and the effect was more significant when 1000 iterations were conducted. The best effect in the experiment was with 4000 iterations. The SSIM of DLDR was 0.9626334, while the SSIM of SR-DLDR was 0.9666951. When PCC + SSIM was used as the evaluation index, the DLDR achieved 0.9727479, and SR-DLDR achieved 0.9770569.

3.2.2. DR comparison between SR-DLDR and VM

In this part of the experimental results, regardless of whether the visual or experimental data are shown in Figure 6 and Table 3, we can see that the alignment results using the DLDR method were better than those of the VM, with an improvement of about 6%. And the SR+DLDR alignment method that uses SR samples had higher alignment accuracy. For more convenient statistical loss function output, the smaller all loss function output in the program design achieved higher image registration accuracy.

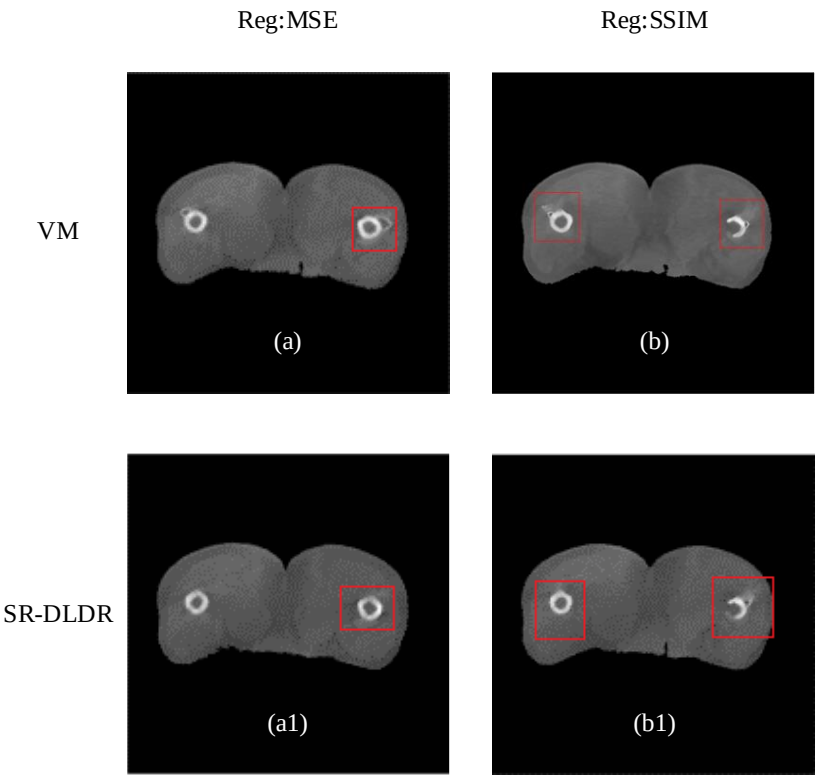

Figure 6. The first row (a) and (b) show the image registration results for the VM model. The second row (a1) and (b1) show the image registration results for the model proposed used in this paper. The first column result used the MSE as the loss function, and the second column used the SSIM as the loss function.

Table 3. Quantitative DR registration comparison between SR-DLDR and VM.

Method	MSE	MI	PCC	SSIM	PCC+SSIM
VM	0.0028766	−	−	0.9056045	−
DLDR	0.0022588	0.5957638	0.9900116	0.9626334	0.9727479
SR-DLDR	0.0022189	0.5979443	0.9900189	0.9666951	0.9770569

4. Discussion

In this paper, we investigate the effect of super-resolution on the registration accuracy. The image quality of the input moving was improved by adding super-resolution before registration, and the registration effect of the image pair with SR processing is compared with that of the image pair without SR processing.

The comparison experiments used two types of mainstream registration methods, i.e., rigid registration, and elastic registration. We compared the registration accuracy of images processed without the SR technique and images processed with the SR technique under the conditions of three rigid registration methods, i.e., rigid transform [49], affine transform [50], and similar transform [51]. The data in Table 1 shows that all three registration accuracies had different degrees of improvement after using SR to preprocess the data, which proves that improving the quality of the input image is helpful to increase the registration accuracy.

For the mainstream registration methods using deep learning, the registration accuracy and generalizability of elastic registration exceed those of rigid registration. We also conducted experiments on elastic registration, because supervised registration requiring the SR processing of images and labeled data is not applicable, and we chose an unsupervised registration algorithm using deep learning methods [40] for testing. The same experimental procedure as the rigid registration method was tested with four loss functions, MSE, MI, PCC and SSIM. Table 2 demonstrates that the images processed with the SR technique are still effective in improving the registration accuracy by up to 4% (under the SSIM loss function). For the feature that images processed by the SR technique have more pixel information, we combined PCC and SSIM loss functions to construct a compound loss function of PCC + SSIM, which can utilize the rich pixel information of the input image while constraining the structure and similarity of the image. From the experimental data, it can be concluded that the newly proposed composite function can further improve the registration accuracy when the input SR-processed image is used, and the improvement reached 5% compared to the image without SR processing.

We demonstrate that the higher the quality of the input image, the higher the registration accuracy, which is applicable under both rigid and elastic registration. However, we have only validated this based on 2D data so far, and have not extended it to 3D data, which is incomplete. All of our experiments were based on clinical pelvic images and it is not possible to confirm whether the method is effective on other sites. In the future, we will further test whether the SR technique can improve the registration accuracy of data in a 3D data environment.

5. Conclusions

In this paper, we have proposed a method to enhance the registered images using SR and

demonstrated the use of a compound loss function PCC + SSIM to enhance the registration effect. We have demonstrated via our experiments that the higher the quality of the image the better the registration. It can also be seen in the results of the experiments that our proposed method can effectively improve the registration accuracy, and that the composite loss function PCC + SSIM used has better performance in registration compared to other loss functions. This method is important for improving the registration accuracy in IGRT and enhancing the treatment results.

Acknowledgments

This work is partially funded by the National Science Foundation for Young Scientists of China (Grant No. 61806060), China. We also acknowledge support from the Basic and Applied Basic Research Foundation of Guangdong Province (2021A1515220140) and the Youth Innovation Project of Sun Yat-sen University Cancer Center (QNYCPY32).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. A. K. Jain, Y. Zhong, S. Lakshmanan, Object matching using deformable templates, *IEEE Trans. Pattern. Anal. Mach. Intell.*, **18** (1996), 267–278. https://doi.org/10.1109/34.485555
2. D. P. Kingma, M, Welling, Auto-Encoding variational bayes, preprint, arXiv:1312/6114.
3. A. Ahmadi, I. Patras, Unsupervised convolutional neural networks for motion estimation, in *IEEE International Conference on Image Processing*, IEEE, (2016), 1629–1633. https://doi.org/10.1109/ICIP.2016.7532634
4. S. Joshi, B. Davis, M. Jomier, G. Gerig, Unbiased diffeomorphic atlas construction for computational anatomy, *NeuroImage*, **23** (2004), S151–S160. https://doi.org/10.1016/j.neuroimage.2004.07.068
5. M. A. Viergevera, J. B. Antoine-Maintzb, S. Kleinc, K. Murphyd, M. Staringe, J. P. W. Pluimfa, A survey of medical image registration-under review, *Med. Image Anal.*, **33** (2016), 140–144. https://doi.org/10.1016/j.media.2016.06.030
6. Y. Fu, N. M. Brown, S. U. Saeed, A. Casamitjana, Y. Hu, DeepReg: a deep learning toolkit for medical image registration, *J Open Source Softw.*, **5** (2020), 2705. https://doi.org/10.21105/joss.02705
7. D. FAIM, A ConvNet Method for Unsupervised 3D Medical Image Registration, in *International Workshop on Machine Learning in Medical Imaging*, Springer, (2019), 646–654. https://doi.org/10.1007/978-3-030-32692-0_74
8. Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, X. Yang, Deep learning in medical image registration: A review, *Phys. Med. Biol.*, **65** (2019), 20TR01. https://doi.org/10.1088/1361-6560/ab843e
9. Y. Fu, S. Liu, H. H. Li, D. Yang, Automatic and hierarchical segmentation of the human skeleton in CT images, *Phys. Med. Biol.*, **62** (2017), 2812–2833. https://doi.org/10.1088/1361-6560/aa6055

10. X. Yang, N. Wu, G. Cheng, Z. Zhou, D. S .Yu, J. J. Beitler, et al., Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy, *Int. J. Radiat. Oncol. Biol. Phys.*, **90** (2014), 1225–1233. https://doi.org/10.1016/j.ijrobp.2014.08.350
11. X. Huang, Y. Zhang, J. Wang, A biomechanical modeling guided simultaneous motion estimation and image reconstruction technique (SMEIR-Bio) for 4D-CBCT reconstruction, *Phys. Med. Biol.*, **63** (2018), 045002. https://doi.org/10.1088/1361-6560/aaa730
12. J. Boda-Heggemann, F. Lohr, F. Wenz, M. Flentje, M. Guckenberger, Cone-Beam CT-Based IGRT, *Strahlenther Onkol.*, **187** (2011), 284–291. https://doi.org/10.1007/s00066-011-2236-4
13. X. Zhen, X. Gu, H. Yan, L. Zhou, X. Jia, S. B. Jiang, CT to Cone-beam CT deformable registration with simultaneous intensity correction, *Phys. Med. Biol.*, **57** (2012), 6807–6826. https://doi.org/10.1088/0031-9155/57/21/6807
14. C. Veiga, J. Mcclelland, S. Moinuddin, A. Lourenço, K. Ricketts, J. Annkah, et al., Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations, *Med. Phys.*, **41** (2014), 031703. https://doi.org/10.1118/1.4864240
15. C. Veiga, G. Janssens, C. L. Teng, T. Baudier, B. K. Teo, First clinical investigation of cone beam computed tomography and deformable registration for adaptive proton therapy for lung cancer, *Int. J. Radiat. Oncol. Biol. Phys.*, **95** (2016), 549–559. https://doi.org/10.1016/j.ijrobp.2016.01.055
16. B. Zhou, Z. Augenfeld, J. Chapiro, S. K. Zhou, C. Liu, J. S. Duncan, Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural CBCT/MR liver segmentation and registration, *Med Image Anal*, **71** (2021), 102041. https://doi.org/10.1016/j.media.2021.102041
17. B. Zhou, C. Liu, J. S. Duncan, Anatomy-Constrained contrastive learning for synthetic segmentation without ground-truth, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2021), 47–56. https://doi.org/10.1007/978-3-030-87193-2_5
18. X. Liang, Y. Jiang, Y. Xie, T. Niu, Quantitative cone-beam CT imaging in radiotherapy: Parallel computation and comprehensive evaluation on the TrueBeam system, *IEEE Access*, **7** (2019), 66226–66233. https://doi.org/10.1109/ACCESS.2019.2902168
19. K. Srinivasan, M. Mohammadi, J. Shepherd, Investigation of effect of reconstruction filters on cone-beam computed tomography image quality, *Australas. Phys. Eng. Sci. Med.*, **37** (2014), 607–614. https://doi.org/10.1007/s13246-014-0291-8
20. L. Ouyang, T. Solberg, J. Wang, Noise reduction in low-dose cone beam CT by incorporating prior volumetric image information, *Med. Phys.*, **39** (2012), 2569–2577. https://doi.org/10.1118/1.3702592
21. W. Kang, M. Patwari, Low Dose Helical CBCT denoising by using domain filtering with deep reinforcement learning, preprint, arXiv:2104/00889.
22. J. Zheng, D. Zhang, K. Huang, Y. Sun, Cone-Beam computed tomography image pretreatment and segmentation, in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, (2018), 25–28. https://doi.org/10.1109/ISCID.2018.00012
23. X. Liang, L. Chen, D. Nguyen, Z. Zhou, X. Gu, et al., Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy, *Phys. Med. Biol.*, **64** (2019), 125002. https://doi.org/10.1088/1361-6560/ab22f9

24. D. C. Hansen, G. Landry, F. Kamp, M. Li, C. Belka, K. Parodi, et al., ScatterNet: A convolutional neural network for cone-beam CT intensity correction, *Med. Phys.*, **45** (2019), 4916–4926. https://doi.org/10.1002/mp.13175
25. G. Wu, M. Kim, Q. Wang, Y. Gao, D. Shen, Unsupervised deep feature learning for deformable registration of MR brain images, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2013), 649–656. https://doi.org/10.1007/978-3-642-40763-5_80
26. M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, N. Komodakis, A deep metric for multimodal registration, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2016). https://doi.org/10.1007/978-3-319-46726-9_2
27. C. Hao, D. Qi, N. Dong, J. Z. Cheng, P. A. Heng, Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2015). https://doi.org/10.1007/978-3-319-24553-9_62
28. S. Miao, Z. J. Wang, L. Rui, A CNN regression approach for Real-Time 2D/3D registration, *IEEE Trans. Med. Imaging*, **35** (2016), 1352–1363. https://doi.org/10.1109/TMI.2016.2521800
29. X. Yang, R. Kwitt, M. Niethammer, Fast rredictive image registration, in *Deep Learning and Data Labeling for Medical Applications*, Springer, (2016), 48–57. https://doi.org/10.1007/978-3-319-46976-8_6
30. D. Detone, T. Malisiewicz, A. Rabinovich, Deep image homography estimation, preprint, arXiv:1606/03798.
31. T. Carvalho, E. Rezende, M. Alves, F. Balieiro, R. B. Sovat, Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN, in *16th IEEE International Conference On Machine Learning And Applications*, IEEE, (2017). https://doi.org/10.1109/ICMLA.2017.00-47
32. J. Krebs, T. Mansi, H. DelingetteZ. Li, F. Ghesu, S. Miao, et al., Robust non-rigid registration through agent-based action learning, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2017), 344–352. https://doi.org/10.1007/978-3-319-66182-7_40
33. J. Zhang, C. Wang, S. Liu, L. Jia, J. Sun, Content-Aware unsupervised deep homography estimation, in *European Conference on Computer Vision*, Springer, (2020), 653–669. https://doi.org/10.1007/978-3-030-58452-8_38
34. M. Bevilacqua, A. Roumy, C. Guillemot, M. L. A. Morel, Super-resolution using neighbor embedding of back-projection residuals, in *2013 18th International Conference on Digital Signal Processing (DSP)*, 2013.07, IEEE, (2013), 1–8. https://doi.org/10.1109/ICDSP.2013.6622796
35. M. Bevilacqua, A. Roumy, C. Guillemot, M. L. A. Morel, Low-Complexity single-image super-resolution based on nonnegative neighbor embedding, in *Proceedings of the British Machine Vision Conference 2012*,(2012), 1–10. https://doi.org/10.5244/C.26.135
36. R. Timofte, V. De, L. V. Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, in *IEEE International Conference on Computer Vision*, IEEE, (2014), 1920–1927. https://doi.org/10.1109/ICCV.2013.241
37. J. Kim, J. K. Lee, K. M. Lee, Accurate image super-resolution using very deep convolutional networks, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2016), 1646–1654. https://doi.org/10.1109/Cvpr.2016.182

38. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409/1556.
39. C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern. Anal. Mach. Intell.*, **38** (2016), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281
40. J. Chen, Y. Li, Y. Du, E. C. Frey, Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks, *Med. Phys.*, **47** (2020). https://doi.org/10.1002/mp.14545
41. R. Sandkühler, C. Jud, S. Andermatt, P. C. Cattin, AirLab: Autograd image registration laboratory, preprint, arXiv:1806/09907.
42. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
43. D. Rueckert, L. I. Sonoda, C. Hayes, D. Hill, M. O. Leach, Nonrigid registration using free-form deformations: application to breast MR images, *IEEE Trans. Med. Imaging*, **18** (2012), 712–721. https://doi.org/10.1109/42.796284
44. T. Fechter, D. Baltas, One shot learning for deformable medical image registration and periodic motion tracking, *IEEE Trans. Med. Imaging*, **39** (2019), 12. https://doi.org/10.1109/TMI.2020.2972616
45. A. V. Dalca, M. Rakic, J. Guttag, M. R. Sabuncu, Learning conditional deformable templates with convolutional networks, in *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Curran Associates Inc., (2019), 806–818.
46. P. Viola, W. Iii, Alignment by maximization of mutual information, *Int. J. Comput. Vision*, **24** (2008), 137v154. https://doi.org/10.1023/A:1007958904918
47. Z. S. Saad, D. R. Glen, G. Chen, M. S. Beauchamp, R. Desai, R. W. Cox, A new method for improving functional-to-structural MRI alignment using local Pearson correlation, *Neuroimage*, **44** (2009), 839–848. https://doi.org/10.1016/j.neuroimage.2008.09.037
48. Z. Shen, X. Han, Z. Xu, M. Niethammer, Networks for joint affine and Non-parametric image registration, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2019), 4224–4233. https://doi.org/10.1109/CVPR.2019.00435
49. M. Staring, S. Klein, J. Pluim, A rigidity penalty term for nonrigid registration, *Med. Phys.*, **34** (2007), 4098–4108. https://doi.org/10.1118/1.2776236
50. L. Hui, P. Du, W. Zhao, L. Zhang, H. Sun, Image registration based on corner detection and affine transformation, in *International Congress on Image & Signal Processing*, IEEE, (2010), 2184–2188. https://doi.org/10.1109/CISP.2010.5647722
51. C. Papazov, D. Burschka, Deformable 3D shape registration based on local similarity transforms, in *Computer Graphics Forum*, Wiley Online Library, (2011), 1493–1502. https://doi.org/10.1111/j.1467-8659.2011.02023.x