

# Respiratory Motion Correction on PET Images Based on 3D Convolutional Neural Network

**Yibo Hou<sup>1</sup>, Jianfeng He<sup>1\*</sup> and Bo She<sup>2\*</sup>**

<sup>1</sup> School of Information Engineering and Automation, Kunming University of Science and Technology  
Kunming, Yunnan 650504 China

[e-mail: 120112624@qq.com, houyibo666@163.com]

<sup>2</sup> PET/CT Center, First People's Hospital of Yunnan  
Kunming, Yunnan 650031 China  
[e-mail: 1471265639@qq.com]

\*Corresponding author: Jianfeng He and Bo She

*Received October 13, 2021; revised May 8, 2021; accepted June 30, 2022;  
published July 31, 2022*

---

## **Abstract**

Motion blur in PET (Positron emission tomography) images induced by respiratory motion will reduce the quality of imaging. Although existing methods have positive performance for respiratory motion correction in medical practice, there are still many aspects that can be improved. In this paper, an improved 3D unsupervised framework, Res-Voxel based on U-Net network was proposed for the motion correction. The Res-Voxel with multiple residual structure may improve the ability of predicting deformation field, and use a smaller convolution kernel to reduce the parameters of the model and decrease the amount of computation required. The proposed is tested on the simulated PET imaging data and the clinical data. Experimental results demonstrate that the proposed achieved Dice indices 93.81%, 81.75% and 75.10% on the simulated geometric phantom data, voxel phantom data and the clinical data respectively. It is demonstrated that the proposed method can improve the registration and correction performance of PET image.

---

**Keywords:** PET images, Respiratory motion correction, Image registration, Deep learning

---

This study has received funding from the National Natural Science Foundation of China, No. 81760306, NO 82160347; Yunnan Key Laboratory of Smart City in Cyberspace Security, No.202105AG070010.

## 1. Introduction

**L**ung cancer is one of the malignant tumors with the fastest increasing morbidity, which seriously threatens human life and health [1]. Positron emission tomography (PET) is an important tool for the early diagnosis of malignant tumors [2]. Patients' respiratory motion will cause the PET imaging artifacts, which will affect the imaging quality. Therefore, it is important for the clinic diagnosis to correct the artifacts.

To solve this problem, He et al. [3] observed the number of photons collected by PET detector ring changes with the position of the respiratory motion, and then proposed a method of system sensitivity gating for the motion correction. Büther et al. [4] used a list-mode data based gated centroid method and spatially segmented tumors by a sinogram for correction of PET images of liver and lung tumors. Although the above approaches can effectively achieve correction for respiratory motion artifacts, some require significant time consumption or do not make efficient use of the information contained in the data.

With the development of deep learning convolutional neural networks (CNN) [5], it has demonstrated its powerful feature extraction capabilities in various medical image processing, however, there are many aspects that can be improved. For instance, He et al. [6] and Ibtehaz et al. [7] proposed residual module and multiple residual module, which can better extract features in deep network model, and effectively solve the problem of gradient disappearance. The spatial transformer network (STN) for image registration was proposed by Jaderberg et al. [8]. STN enables spatial operations on data in the network, and spatial transformer according to the image to be registered. Wu et al. [9] proposed a stacked autoencoder (SAE) to identify the inherent features in the fixed image to register the image. Sokooti et al. [10] proposed a multi-scale registration framework that integrates more global information. Subsequently, Vos et al. [11] and Li et al. [12] successively proposed a self-supervised full convolutional network (FCN) registration method, but these methods are transformations in a small range without focusing on the entire image. Balakrishnan et al. [13] proposed a fast unsupervised medical image registration method, VoxelMorph, which learns a parameterized registration function using a convolutional neural network (CNN), however, this method has some defects in the registration of complex medical images. A commonly used registration method is AntsSynQuick [14], which uses geometric or variational methods for registration, but it is very computationally intensive. Kuang et al. [15] proposed FAIM which extends VoxelMorph by improving the accuracy of results with penalty loss on negative Jacobian determinants, only from the aspect of image folding. Hu et al. [16] designed a two-stream architecture, Dual-PRNet which extends VoxelMorph by computing multi-scale registration fields from convolutional feature pyramids, but the correlation between images is ignored.

Although the exiting mentioned above methods for the motion correction have obviously improvements, there are still some limitations. In this paper, we propose a framework as Res-Voxel, for respiratory motion correction of PET imaging by incorporating multiple residual blocks into VoxelMorph network, which may achieve a better performance of the motion correction.

## 2. Method

### 2.1 Structure of the Correction System

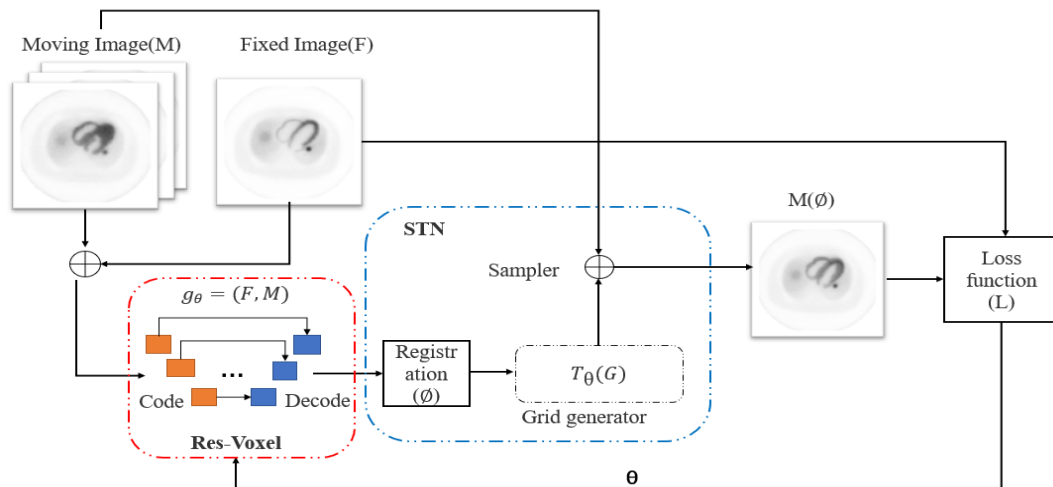
In general, most of registration algorithms are to warp and transform the voxels on the moving image to correspond the voxels on the fixed image[17]:

$$\hat{\Phi} = \arg \min_{\phi} \mathcal{L}(F, M, \phi) \quad (1)$$

$$\mathcal{L}(F, M, \phi) = \mathcal{Lsim}(F, M(\phi)) + \lambda \mathcal{Lsmooth}(\phi) \quad (2)$$

Where  $F$  is fixed image,  $M$  is moving image,  $\phi$  is displacement vector field (DVF) which is the offset of each voxel in  $F$  to the corresponding voxel in  $M$ ,  $M(\phi)$  is the moving image of  $M$  after distortion transformation,  $\mathcal{Lsim}(\cdot)$  is used to measure the similarity between  $F$  and  $M(\phi)$ ,  $\mathcal{Lsmooth}(\cdot)$  is the regularization acting on  $\phi$ , and  $\lambda$  is the regularization parameter.

As shown in Fig. 1, the structure of network for image registration is given. This paper uses a convolutional neural network to model the function  $g_{\theta}(F, M) = \phi$ . In the experiment,  $M$  and  $F$  are used as the input of Res-Voxel, both are grayscale images. The registration domain  $\phi$  is learned through the parameter  $\theta$ . The output  $\phi$  of Res-Voxel is used as the input of the STN that the spatial transformer function is used to transform  $M(p)$  into  $M(\phi(p))$ , making  $M(\phi(p))$  a structural position similar to  $F(p)$ . Finally, the difference between  $M(\phi(p))$  and  $F(p)$  is calculated by the Loss Function, thus the parameter  $\theta$  is updated.



**Fig. 1.** Overall registration framework. The region of red dash line is the Res-Voxel model, which is used to predict the deformation field. The region of blue dash line is the spatial transformer network, which is used to obtain the corrected image.

### 2.2 Res-Voxel

The registration method based on deep learning is used for the deformation field learning through the network based on U-net [18][19], which is a medical image registration framework VoxelMorph, and has achieved results on brain MR image registration in three dimensions [13]. In PET image artifact correction, we focus on the correspondence between the artifact/moving image of the lung and the fixed image. In most cases, the artifact generated by respiratory motion has irregular and complex characteristics, and the U-net has

poor prediction effect on its deformation field. Szegedy et al. [20] proposed Inception architecture. Inception blocks utilized convolutional layers of varying kernel sizes in parallel, to inspect the points of interest in images from different scales. These perceptions, obtained at different scales, were combined together and passed on deeper into the network, which the feature extraction ability of the network is improved. However, the network has still a limitation that the  $3\times 3$  and  $5\times 5$  convolution operations increase the computational burden of the network as shown in Fig. 2(a). We thus use the receptive field size of two  $3\times 3$  convolution kernels to take place of a  $5\times 5$  convolution kernel do decrease the computational burden as shown in Fig. 2(b). Specifically, for an input  $5\times 5$  image as shown in Fig. 2(a), an input image feature map is obtained after a  $5\times 5$  convolution operation that requires the total number of parameters being  $(5\times 5)\times\text{channels}$ . In contrast, an input image feature map is obtained after a  $3\times 3$  convolution operation that requires the total number of parameters being  $(3\times 3)\times\text{channels}$ . As result, two  $3\times 3$  ( $2\times 3\times 3$ ) convolution kernels obviously require less computation as shown in Fig. 2(b), the total number of parameters being  $2\times(3\times 3)\times\text{channels}$  compare to one  $5\times 5$  convolution core, which the parameters reduce by about 30%. Fig. 3 shows the amount of computation of two convolution kernels on images of different sizes. The complexity of both is  $O(n^2)$ . For an image of size  $n$ , the number of convolutions of one  $5\times 5$  convolution kernel is  $(n-4)^2$ , and the amount of computation is  $5\times 5\times(n-4)^2$ ; the number of convolutions of two  $3\times 3$  convolution kernels is  $(n-2)^2+(n-2)^2$ , and the amount of computation is  $2\times(3\times 3)\times[(n-2)^2+(n-2)^2]$ . When the image side length is greater than 10, the computational complexity of two  $3\times 3$  convolutions is smaller than that of one  $5\times 5$  convolution.

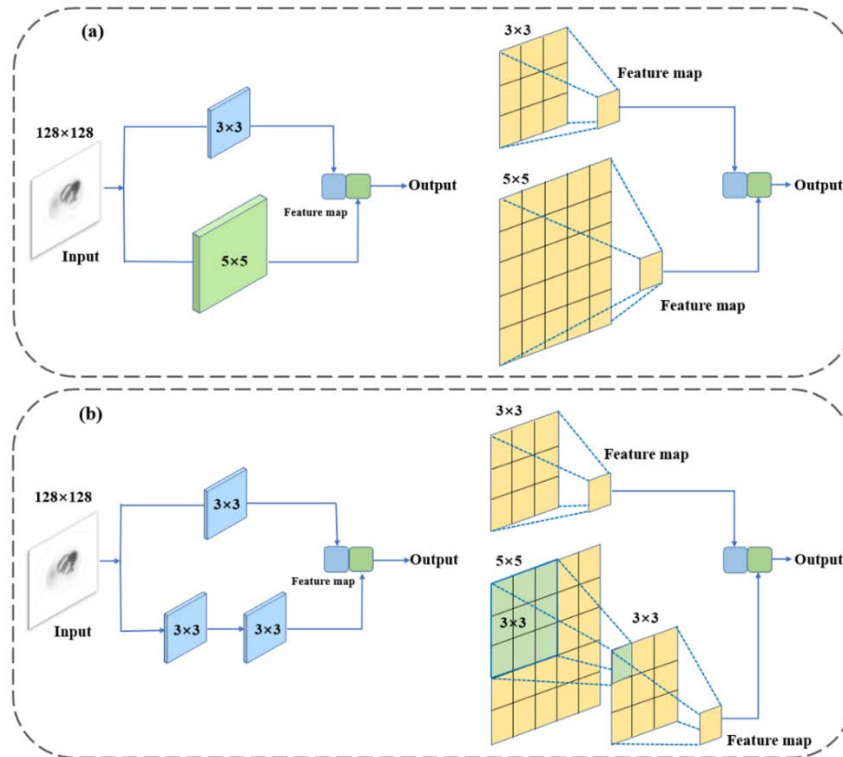


Fig. 2. The two  $3\times 3$  taking place of  $5\times 5$  convolution operations decrease the computational burden of the network

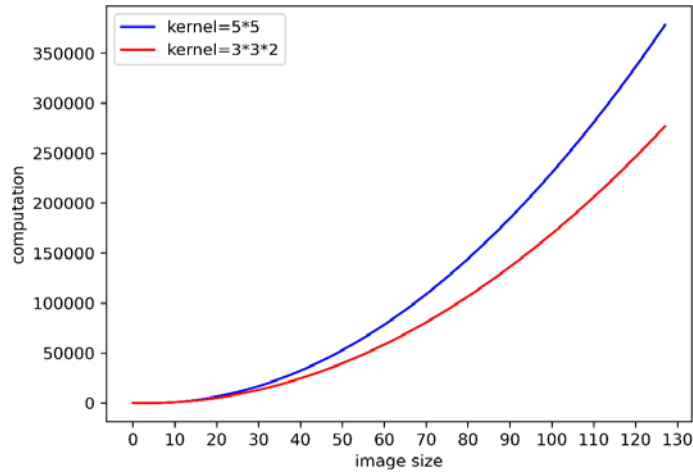


Fig. 3. Computation of different convolution kernels VS images of different sizes

Furthermore, we also supplement relevant experimental results to support that the proposed model requires fewer parameters. As shown in Table 1, we simply use the geometric phantom dataset to compare the number of parameters and the training time of our method with the baseline network under the running of 1000 iterations. We can see the proposed method Res-Voxel reduces the number of parameters by about 7%, so we choose a structure similar to Fig. 2b to construct the registration network.

Table 1. Comparison of parameter results under 1000 iterations

Model	Parameters	Time(s)
VoxelMorph (baseline)	2203392	973
Res-Voxel (our method)	2048768	778

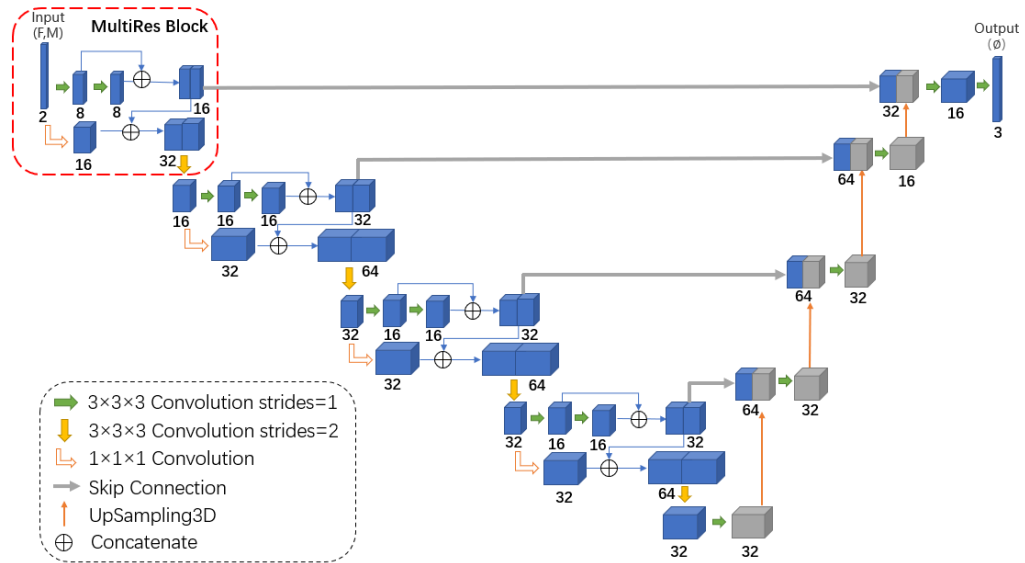


Fig. 4. The structure of Res-Voxel network as shown the region of red dash line in Fig. 1

We add multiple residual blocks at each stage to enable it to extract spatial features from different scales. The structure is shown in Fig. 4. The fixed image ( $F$ ) and the moving image ( $M$ ) are connected into a 2-channel 3D image in the channel dimension as the input of the network. The input is of size  $128 \times 128 \times 32 \times 2$ , and the number of channels in each block is marked in Fig. 4, 3D convolution is used in both encoding and decoding stages, the size of the convolution kernel is  $3 \times 3 \times 3$ , and each convolution layer is followed by an Leaky ReLU as activation function. The convolution step size (stride=2) is used to halve the spatial dimension until the smallest layer is reached. In the decoding stage, convolution, upsampling3D and skip connection are alternately used, and finally gain the registration domain  $\phi$  as output. In experiment, the output size is  $128 \times 128 \times 32 \times 3$ .

### 2.2.1 Multiple Residual Block

Based on the structure and advantages described in Fig. 2, we built this module. By adding a residual connection such as path  $a$ , as well as to introduce  $1 \times 1 \times 1$  convolutional layers such as path  $b$ , we call this arrangement a Multiple residual block, as shown in Fig. 5:

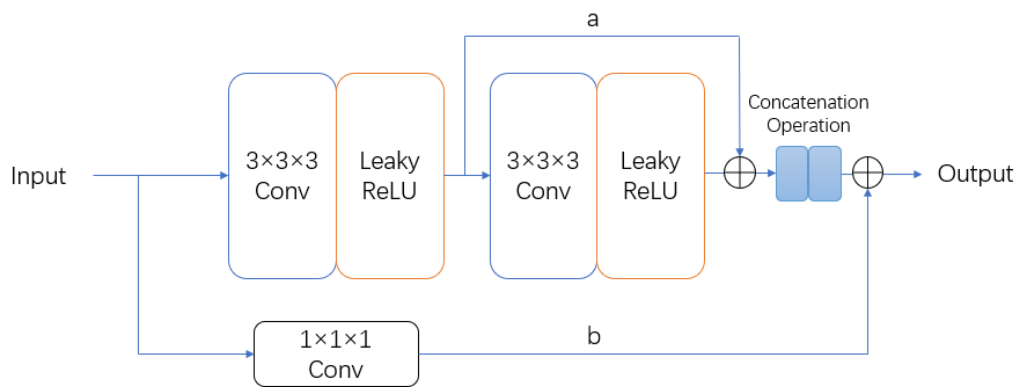
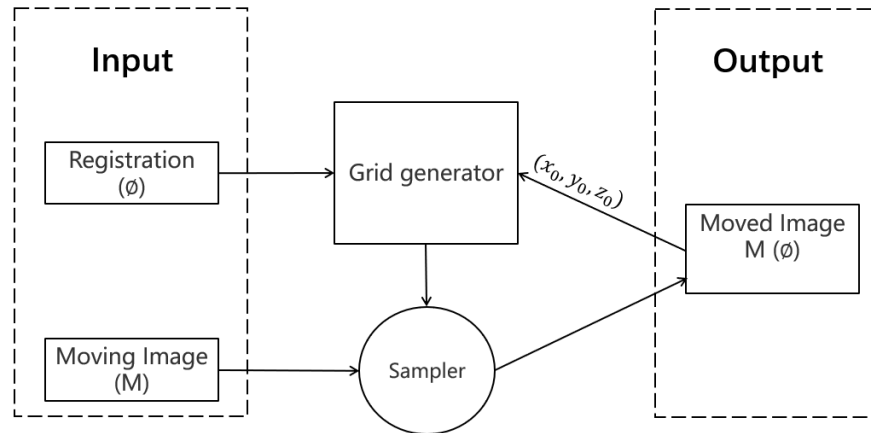


Fig. 5. Multiple residual block as shown the region of red dash line in Fig. 3.

In the block, the input image passes through two consecutive  $3 \times 3 \times 3$  convolution, splicing with previous one result, and then spliced with the  $1 \times 1 \times 1$  convolutional image, and finally the dimensionality reduction operation is performed through a  $3 \times 3 \times 3$  convolution with a step size of 2. As this structure, image features are extracted through a  $3 \times 3 \times 3$  convolution operation, and some additional spatial information can be obtained by an  $1 \times 1 \times 1$  convolution operation. The entire block can effectively avoid the problem of excessive parameters, and the image features are well extracted at the same time.

### 2.3 Spatial Transformer Network

The spatial transformer network [8] designed in this paper is used to calculate  $M(\phi)$ . The structure is shown in Fig. 6:



**Fig. 6.** Spatial Transformer Networks as shown the region of blue dash line in Fig. 1

The network is composed of input, Grid generator, Sampler and output. The input is composed of moving image  $M$  and registration  $\phi$ , and the output is moved image  $M(\phi)$ . The Grid generator is used to calculate the coordinates in  $M$ , according to the coordinates in  $M(\phi)$  and the parameter  $\phi$ , because of each voxel value in  $M(\phi)$  must be extracted from  $M$  and then filled. The Sampler is used to fill  $M(\phi)$  according to the coordinates obtained in the Grid generator and the corresponding voxel value in  $M$ . For each voxel  $p$ , the position of the voxel  $\phi(p)$  calculated in  $M$  may be a decimal, while the voxel value should be integer, so that linear interpolation is performed on the eight adjacent voxel in the experiment, which is as follows:

$$M(\phi(p)) = \sum_{q \in Z(\phi(p))} M(q) \prod_{i \in \{x,y,z\}} (1 - |\phi_i(p) - q_i|) \quad (3)$$

where the  $Z(\phi(p))$  is the adjacent voxel of  $\phi(p)$ , and  $\phi_i(p)$  represents the coordinates of the corresponding point found in  $M$  for the  $i$ -th point in  $M(\phi)$ .

## 2.4 Loss Function and Evaluation Indicators

In the experiment, AdaBound [21] is used as the optimization algorithm of the model. The parameters are set by the learning rate with 0.001, the exponential attenuation rate of the first-order moment estimation with 0.9, the decay rate of the second-order moment exponent with 0.99, and the disturbance value with 10E-8.

A loss function  $L$  is required in the optimization algorithm AdaBound, and mean square error (MSE) is used. The MSE function curve is continuous and conducive to derivation, which is convenient for the use of the AdaBound optimization algorithm. Moreover, as the error decreases, the gradient also decreases, which is conducive to convergence and makes the result more stable. The degree of data transformation is evaluated by the expected value of the square of the difference between the estimated value of the parameter and the true value of the parameter. In this paper, it is expressed as the difference between the corresponding voxels of the image  $M(\phi)$  after distortion and the fixed image  $F$ .

Since a large deformation may occur after the linear interpolation distortion transformation  $M(\phi)$ , a bending energy penalty is added to make it more smooth, which is:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p_i)\|^2 \quad (4)$$

The final loss function is:

$$L(F, M, \phi) = MSE(F, M, \phi) + L_{smooth}(\phi) \quad (5)$$

We use Dice similarity coefficient (Dice) [22], Intersection Over Union (IOU) [23] and Hausdorff Distance (HD) [24] to evaluate the network performance. Both Dice and IOU are used to measure the similarity of two sets. In experiments, they are used to evaluate the result of registration and correction. The larger the value of Dice and IOU are, the more similar is between the correction image and the fixed image. The HD is used to measure the boundaries. The smaller the HD value, the smaller the difference is between the correction image and the fixed image.

### 3. Experiments

#### 3.1 Dataset

In this paper, a total of 120 geometric phantoms of small cylinders under different frequencies and motion directions are simulated for simple respiratory motion by GATE (GEANT4 Application Tomographic Emission) [25]. The respiratory motion amplitude of each simulation is between -3cm and 3cm, the motion period is controlled between 4s and 5s, and the radioactivity is 3000KBq, the size of the image is 128×128×16, and the resolution of the image is 3.125mm×3.125mm×4.25mm.

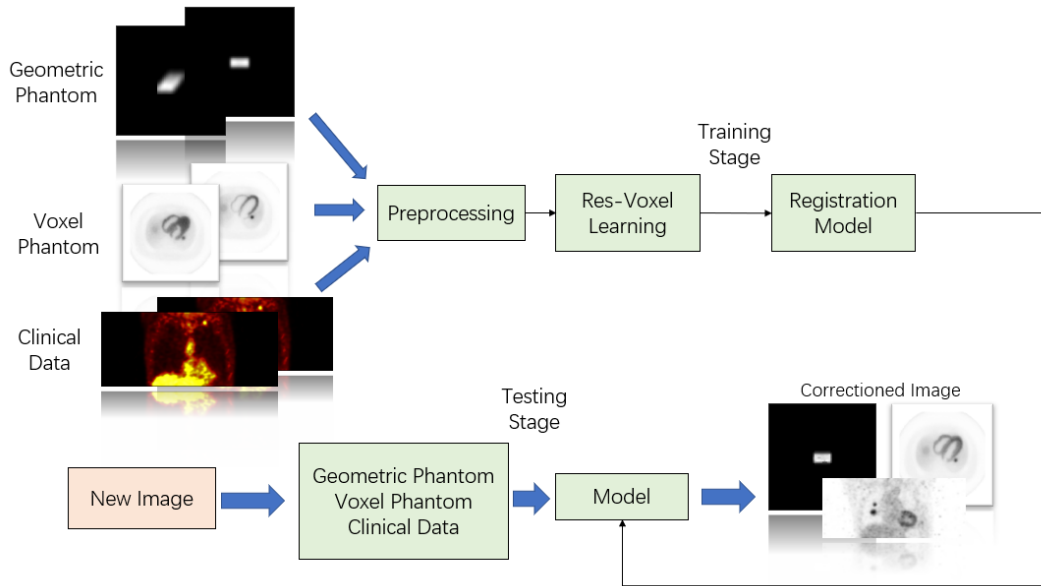
In addition, 100 cases of voxel phantoms with different respiratory frequencies and different motion amplitudes are simulated by NCAT (NURBs [Non Uniform Rational B-Splines] Cardiac Torso) [26] that simulates the dynamic phantom of various organs of the body under respiratory motion. The respiratory motion amplitude of each model is set between 1cm and 3cm, the respiratory cycle is 5s, and the radioactive activity of the lungs is set to 3KBq, the radioactivity of the tumor is set to 20KBq, the size of the simulated image is 128×128×32, and the resolution of the image is 3.125mm×3.125mm×4.25mm. Specific parameter settings are shown in Table 2.

**Table 2.** Phantom parameter setting

	<b>Amplitude (cm)</b>	<b>Period (s)</b>	<b>Radioactivity (KBq)</b>	<b>Image Size</b>	<b>Resolution (mm)</b>
geometric phantom	0~5.2	4~5	3000	128×128×16	3.125×3.125 ×4.25
voxel phantom	1~3	5	Lung (3) Tumor (20)	128×128×32	3.125×3.125 ×4.25

Fig. 7 shows the experimental process. The raw data generated by the GATE and NCAT simulation is in the file of root format. In the experiment, STIR [27] and OSEM algorithm are used to reconstruct the data. Then, the obtained data is cropped, zero-averaged and normalized, so that the voxel value of the image is in the range of 0~1, where the mean value of the data becomes 0, and the standard deviation is 1, which can make the gradient descent more effective and accelerate the convergence of the model. The geometric phantoms are divided into training set and test set according to the ratio of 8:2 (96 cases for the training set and 24 cases for the test set). The voxel phantoms are divided into training set and test set according to the ratio of 9:1 (90 cases for the training set and 10 cases for the test set).





**Fig. 7.** Experimental process. The experimental datasets are divided into geometric phantoms, voxel phantoms and clinical data. In the training stage, the corresponding registration model is trained through the Res-Voxel model according to different datasets. In the test phase, input a new image to the model and output the corrected image.

### 3.2 Computation Environments

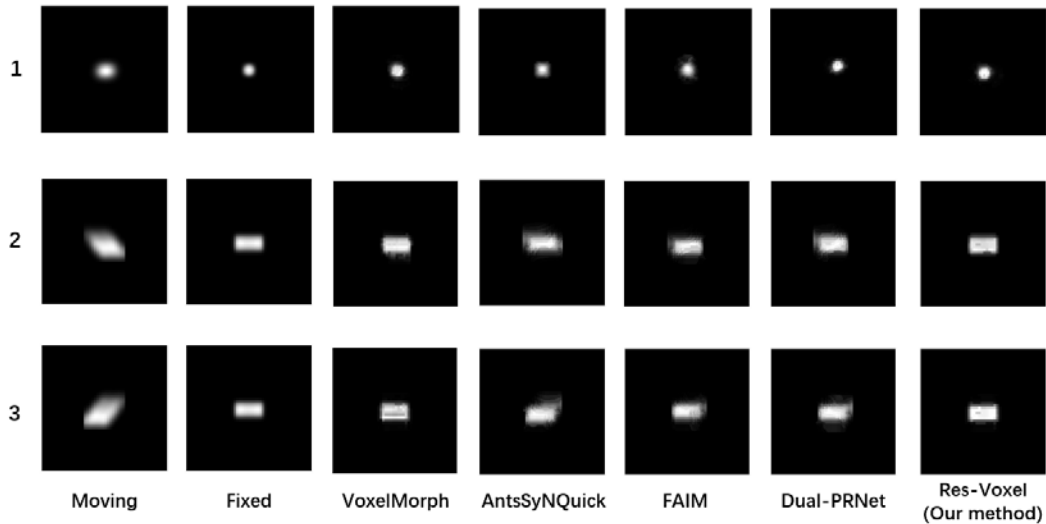
This paper uses Pycharm as the compiler. The programming language is python3.6, the experimental framework is TensorFlow, and the computer hardware configuration is Intel(R) Core(TM) i7-9700KF CPU@3.60 GHz, 64GB RAM, NVIDIA GeForce RTX 2080 Ti 11G graphic memory, 64 Bit Windows10 operating system.

## 4. Results

### 4.1 Geometric phantom correction

Geometric phantom was used to simulate simple respiration registration experiment. In the training phase, we input a fixed image and a series of moving images into the network, as shown in Fig. 1. In the test phase, we input a fixed image and a moving image into the trained registration model for end-to-end registration and correction. The following is the evaluation and analysis of the results of the 2D slices of the transverse, coronal and sagittal views of the geometric phantom.

Fig. 8 is a result of artifact correction for one of the samples. 1 to 3 lines are transverse, coronal and sagittal images respectively. The columns from left to right are moving images with artifacts, fixed images without artifacts. The experimental results are compared with the traditional VoxelMorph [13] and AntsSynQuick [14] methods and the state-of-the-art FAIM [15] and Dual-PRNet [16] methods.



**Fig. 8.** Correction results of geometry phantom

It can be seen from the **Fig. 8** that there is a great artifact effect around the moving image in the first column, the small cylinder becomes blurred, and all methods after registration correction have almost no artifacts. VoxelMorph can correct most artifacts well, and Res-Voxel has a better correction effect on the periphery of the small cylinder, which also shows the effectiveness of the proposed.

In the experiment, Dice [22], IOU [23] and HD [24] are used to quantitatively analyze the experimental results, compare and explain the images before and after registration as shown in **Table 3**:

**Table 3.** PET/CT geometry phantom evaluation results

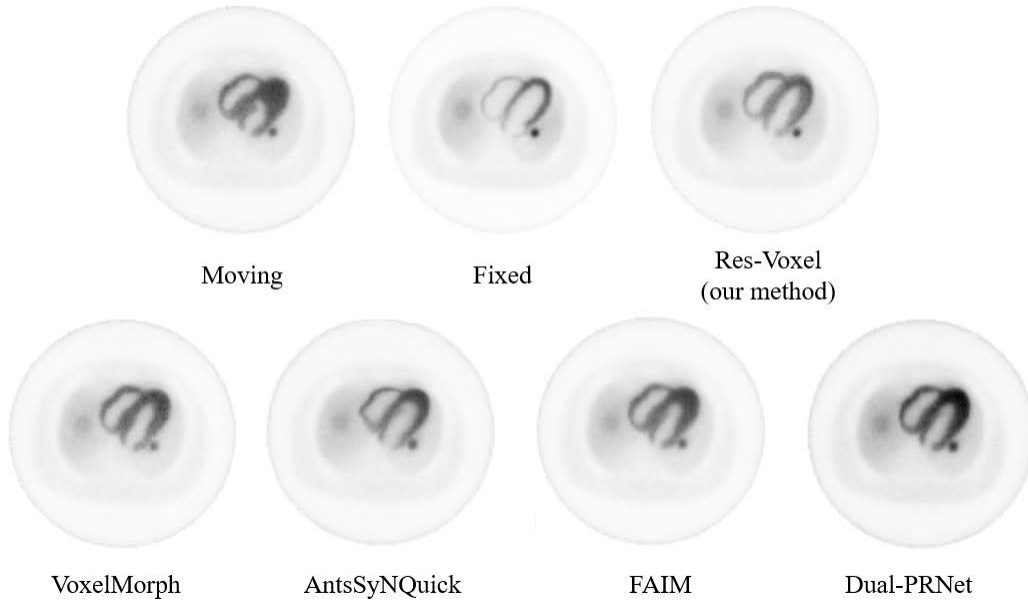
		<b>Dice (%)</b>	<b>IOU (%)</b>	<b>HD</b>
Transverse	Moving	60.32	43.18	9.32
	VoxelMorph (baseline)	84.04	72.48	5.91
	AntsSyNQuick	85.25	73.30	3.82
	FAIM	67.45	50.28	8.74
	Dual-PRNet	86.76	76.26	<b>3.16</b>
	Res-Voxel (our method)	<b>89.04</b>	<b>80.25</b>	5.91
Coronal	Moving	60.79	43.66	11.04
	VoxelMorph (baseline)	92.91	87.51	8.33
	AntsSyNQuick	82.25	76.55	7.58

	FAIM	93.10	85.07	4.00
	Dual-PRNet	89.91	80.53	7.87
	Res-Voxel (our method)	<b>93.81</b>	<b>88.35</b>	<b>5.83</b>
Sagittal	Moving	65.05	48.21	10.19
	VoxelMorph (baseline)	89.20	80.51	8.36
	AntsSyNQuick	83.35	77.88	6.29
	FAIM	86.13	79.45	6.09
	Dual-PRNet	87.45	80.20	7.58
	Res-Voxel (our method)	<b>90.13</b>	<b>82.04</b>	<b>5.83</b>

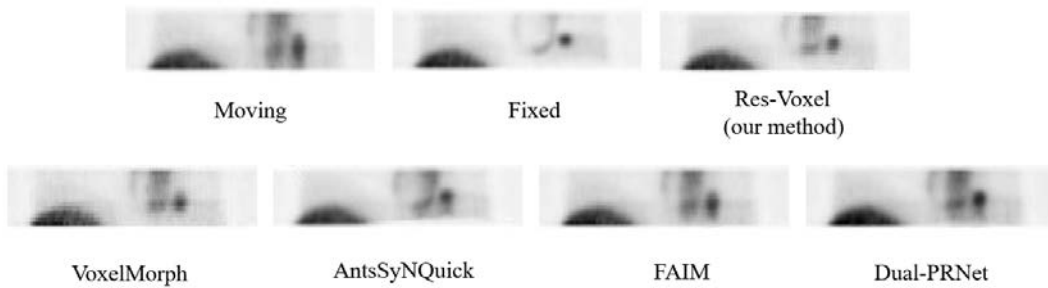
It can be concluded that the evaluation value after registration has been greatly improved compared with the value before registration, especially in the Dice and IOU indicators of the similarity measurement. Res-Voxel achieved 89.04%, 93.81% and 90.13% of the Dice assessment results in the transverse, coronal and sagittal planes, respectively. Res-Voxel consistently outperforms the other three methods. The difference with the fixed image is smaller, and the Dice value of the coronal plane is increased by about 0.90% compared with VoxelMorph.

#### 4.2 Voxel phantom correction

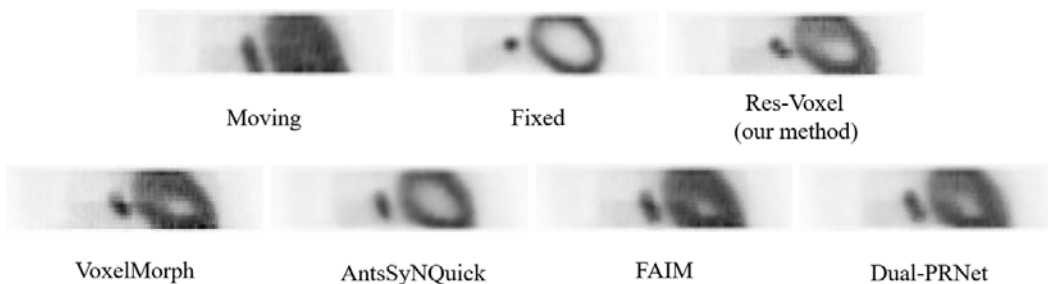
In order that further verify the effectiveness of the Res-Voxel framework, the lung voxel phantom data is simulated on the NCAT from the basis of the internal tissues and organs of the real human body data. In the registration experiment of the voxel phantom, we divide a breathing cycle of 5s into 10 frames, each frame is 0.5s, and then select one of the images (the image of the sixth frame in this experiment) as a fixed image, input other frame images as moving images into the network model for registration correction, and finally merge the obtained results to realize the artifact correction of breathing movement. The result of registration correction is shown in **Fig. 9-Fig. 11**:



**Fig. 9.** Transverse correction results of voxel phantom



**Fig. 10.** Coronal correction results of voxel phantom



**Fig. 11.** Sagittal correction results of voxel phantom

It can be seen from the **Fig. 9- Fig. 11** that the moving image has artifacts both around the lung and around the tumor, making it difficult to observe its exact position and structure, compared with the fixed image. The image quality after registration correction is significantly improved, and Res-Voxel has a better correction effect on tumor artifacts than

other methods. Especially in the coronal and sagittal planes, it has a better correction effect on the edges of organs and tumors. Specific related evaluation results are shown in [Table 4](#).

**Table 4.** PET/CT voxel phantom evaluation results

		Dice (%)	IOU (%)	HD
Transverse	Moving	78.47	64.57	5.31
	VoxelMorph (baseline)	80.55	67.86	5.14
	AntsSyNQuick	81.62	68.94	5.48
	FAIM	78.98	65.26	5.29
	Dual-PRNet	80.85	68.76	6.11
	Res-Voxel (our method)	<b>81.99</b>	<b>69.47</b>	<b>5.03</b>
Coronal	Moving	67.54	50.98	5.83
	VoxelMorph (baseline)	80.60	67.50	4.47
	AntsSyNQuick	76.57	62.03	10.04
	FAIM	79.39	65.82	5.83
	Dual-PRNet	80.73	67.85	6.08
	Res-Voxel (our method)	<b>81.75</b>	<b>68.87</b>	<b>4.35</b>
Sagittal	Moving	77.42	63.15	5.38
	VoxelMorph (baseline)	81.91	69.36	5.19
	AntsSyNQuick	81.70	69.06	5.65
	FAIM	80.05	66.74	5.47
	Dual-PRNet	81.32	68.52	5.65
	Res-Voxel (our method)	<b>83.25</b>	<b>71.31</b>	<b>5.29</b>

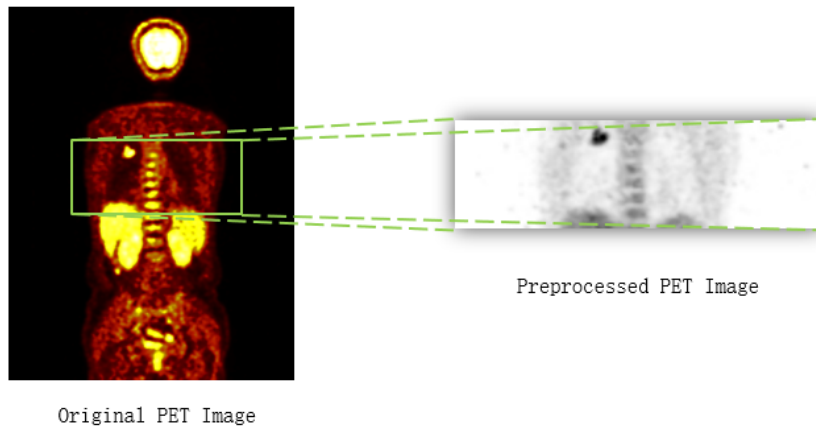
It can be drawn from the [Table 4](#) that the evaluation results after correction are greatly improved compared with those before correction. The Dice indices of Res-Voxel in the transverse, coronal, and sagittal planes are 71.91%, 81.75%, and 83.25%, respectively, which are closer to the fixed image, and are approximately 1.02% higher than Dual-PRNet with better result on the obvious coronal plane. It also illustrates the effectiveness of the proposed approach.

### 4.3 Clinical data correction

In this section, relevant experiments are carried out on clinical data, and the experimental results are evaluated and analyzed, so as to verify the feasibility of the proposed method for the correction of respiratory motion artifacts.

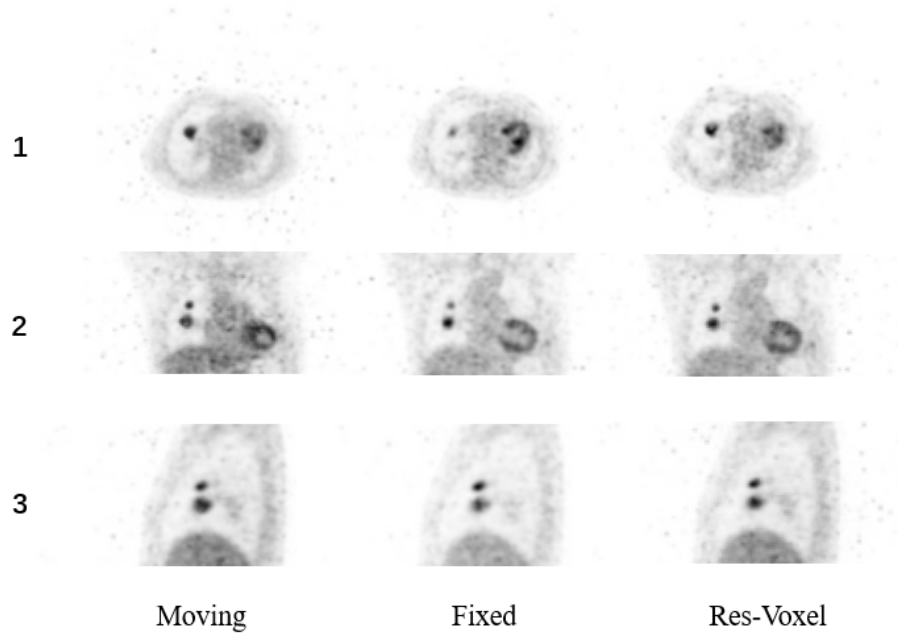
The clinical data were collected from the PET / CT center of the first people's Hospital of Yunnan Province, China. All patients were examined on Philips Ingenuity TF PET / CT scanner, Philips Intellispace portal v7 0.4.20175 for post-treatment. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the First People's Hospital of Yunnan Province (No. 2017YYLH035). Prior informed consent to participate was obtained from all participants.

A total of eight patient data were used in this experiment, and the original data size was  $144 \times 144 \times 234$  whole body PET images. The original PET image is cropped to a size of  $128 \times 128 \times 60$ ; through zero mean and normalization, the training speed of the network is improved; then it is linearly stretched to improve the contrast of the PET image and highlight the important information in the image. The preprocessed PET image is shown in [Fig. 12](#).



**Fig. 12.** PET image comparison before and after preprocessing

In this experiment, the PREVIEW image is selected as the fixed image, and the PREVIEW image is the predicted image generated by the PET/CT scanner using the gating method during the reconstruction process. Therefore, the position and shape of the corresponding organ tissue and tumor in the obtained image are relatively fixed, but the data utilization rate is low, and the signal-to-noise ratio of the obtained image is low, as shown in the Fixed image in [Fig. 13](#).



**Fig. 13.** Correction results of clinical data

From **Fig. 13** compared with the fixed image, the moving image of the heart part has become blurred, and there is no obvious boundary, and the volume of the tumor is also larger than that of the fixed image because of the respiratory motion, and the corrected tumor is slightly reduced. Smaller, closer to a fixed image, and organs such as the heart become slightly clearer. In order to analyze the experimental results more objectively, the experiments are combined with the relevant evaluation index data to carry out relevant quantitative analysis of the experimental results. The specific results are shown in **Table 5**.

**Table 5.** Clinical data evaluation results

		<b>Dice (%)</b>	<b>IOU (%)</b>	<b>HD</b>
Transverse	Moving	70.88	63.21	8.34
	Res-Voxel (our method)	<b>77.86</b>	<b>66.37</b>	<b>7.11</b>
Coronal	Moving	67.78	53.57	6.89
	Res-Voxel (our method)	<b>75.10</b>	<b>57.23</b>	<b>5.63</b>
Sagittal	Moving	62.45	53.23	6.56
	Res-Voxel (our method)	<b>74.66</b>	<b>55.49</b>	<b>6.38</b>

As can be seen from **Table 5** the similarity Dice values obtained by the registration results of the method in this paper in the transverse, coronal and sagittal planes are 77.86%, 75.10% and 74.66%, respectively. Increased by 6.98%, 7.32% and 12.21%. Combined with

the results of subjective evaluation and quantitative analysis, it shows that the registration correction method in this paper has a certain correction effect on clinical PET respiratory motion images.

## 5. Conclusion

For intricate PET images suffering from noises and lack of clear boundaries, we proposed an unsupervised framework, Res-Voxel, for PET image artifact correction. The network integrates multiple residual blocks and uses multiple smaller convolution operations to improve the ability of model to predict the deformation field, while reducing the parameters of the model and decreasing the amount of computation. We performed experimental verification on the simulated geometric phantom data, voxel phantom data and clinical data, which shows that the proposed approach can effectively reduce the respiratory motion artifacts in PET images and improve the image quality.

## References

- [1] Siegel R L, Miller K D, Goding Sauer A, et al, "Colorectal cancer statistics, 2020," *CA: a cancer journal for clinicians*, vol. 70, no. 3, pp. 145-164, 2020. [Article \(CrossRef Link\)](#)
- [2] Cherry S R, Dahlbom M, "PET: physics, instrumentation, and scanners," in *PET*, New York, NY: Springer, 2006, pp. 1-117. [Article \(CrossRef Link\)](#)
- [3] He J, O'Keefe G J, Gong S J, et al, "A novel method for respiratory motion gated with geometric sensitivity of the scanner in 3D PET," *IEEE transactions on nuclear science*, vol. 55, no. 5, pp. 2557-2565, 2008. [Article \(CrossRef Link\)](#)
- [4] Florian Büther, Iris Ernst, Mohammad Dawood, Peter Kraxner, Michael Schäfers, Otmar Schober, Klaus P. Schäfers, "Detection of respiratory tumour motion using intrinsic list mode-driven gating in positron emission tomography," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, pp. 2315-2327, 2010. [Article \(CrossRef Link\)](#)
- [5] LeCun Y, Bengio Y, Hinton G, "Deep learning," *nature*, vol. 521, pp. 436-444, 2015. [Article \(CrossRef Link\)](#)
- [6] He K, Zhang X, Ren S, et al, "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016. [Article \(CrossRef Link\)](#)
- [7] Nabil Ibtehaz, M. Sohel Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74-87, 2020. [Article \(CrossRef Link\)](#)
- [8] Jaderberg M, Simonyan K, Zisserman A, et al, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, pp. 2017-2025, 2015. [Article \(CrossRef Link\)](#)
- [9] Wu Guorong, Kim Minjeong, Wang Qian, Munsell Brent C, Shen Dinggang, "Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning," *IEEE transactions on biomedical engineering*, vol. 63, no. 7, pp. 1505-1516, 2016. [Article \(CrossRef Link\)](#)
- [10] Sokooti H, De Vos B, Berendsen F, et al, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 232-239, 2017. [Article \(CrossRef Link\)](#)
- [11] de Vos B D, Berendsen F F, Viergever M A, et al, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Proc. of International Workshop on Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 204-212, 2017. [Article \(CrossRef Link\)](#)
- [12] Li H, Fan Y, "Non-rigid image registration using fully convolutional networks with deep self-supervision," *arXiv preprint arXiv:1709.00799*, 2017. [Article \(CrossRef Link\)](#)



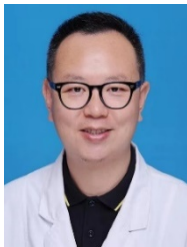
- [13] Balakrishnan G, Zhao A, Sabuncu M R, et al, "An unsupervised learning model for deformable medical image registration," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 9252-9260, 2018. [Article \(CrossRef Link\)](#)
- [14] Avants B B, Tustison N J, Song G, et al, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033-2044, 2011. [Article \(CrossRef Link\)](#)
- [15] Kuang D, Schmah T, "Faim—a convnet method for unsupervised 3d medical image registration," in *Proc. of International Workshop on Machine Learning in Medical Imaging*, Springer, Cham, pp. 646-654, 2019. [Article \(CrossRef Link\)](#)
- [16] Hu X, Kang M, Huang W, et al, "Dual-stream pyramid registration network," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 382-390, 2019. [Article \(CrossRef Link\)](#)
- [17] M. Faisal Beg, Michael I. Miller, Alain Trouvé, Laurent Younes, "Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139-157, 2005. [Article \(CrossRef Link\)](#)
- [18] Isola P, Zhu J Y, Zhou T, et al, "Image-to-image translation with conditional adversarial networks," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134, 2017. [Article \(CrossRef Link\)](#)
- [19] Ronneberger O, Fischer P, Brox T, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. of International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015. [Article \(CrossRef Link\)](#)
- [20] Szegedy C, Liu W, Jia Y, et al, "Going deeper with convolutions," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015. [Article \(CrossRef Link\)](#)
- [21] Luo L, Xiong Y, Liu Y, et al, "Adaptive gradient methods with dynamic bound of learning rate," *arXiv preprint arXiv:1902.09843*, 2019. [Article \(CrossRef Link\)](#)
- [22] Dice L R, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297-302, 1945. [Article \(CrossRef Link\)](#)
- [23] Rahman M A, Wang Y, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proc. of International symposium on visual computing*, pp. 234-244, 2016. [Article \(CrossRef Link\)](#)
- [24] Blumberg H, "Hausdorff's Grundzüge der Mengenlehre," *Bulletin of the American Mathematical Society*, vol. 27, no. 3, pp. 116-129, 1920. [Article \(CrossRef Link\)](#)
- [25] Jan S, Santin G, Strul D, et al, "GATE: a simulation toolkit for PET and SPECT," *Physics in Medicine & Biology*, vol. 49, no. 19, p. 4543, 2004. [Article \(CrossRef Link\)](#)
- [26] Segars W P, "Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom," *The University of North Carolina at Chapel Hill*, 2001.
- [27] Fuster B M, Falcon C, Tsoumpas C, et al, "Integration of advanced 3D SPECT modeling into the open-source STIR framework," *Medical physics*, vol. 40, no. 9, 2013. [Article \(CrossRef Link\)](#)



**Yibo Hou** received the B.S. degree in software engineering from Shanxi University, Shanxi, China, in 2018. He is a M.S. candidate in software engineering with the School of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, China. His research interests include computer vision, deep learning, and medical image processing.



**Jianfeng He** received the Ph.D. degree from Royal Melbourne Institute of Technology University in 2008. He is a Professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, China. His research interests include computer vision, data mining, artificial intelligence and medical image processing.



**Bo She** received the M.S from Kunming Medical University in medical imaging. He is deputy chief physician in the PET-CT center of the first people's Hospital of Yunnan, engages in medical imaging and PET-CT diagnosis. In 2014, he went to the PET-CT center of the General Hospital of the Chinese People's Liberation Army (Beijing 301 Hospital) for further study, and in 2017, he visited the PET-CT center of the Affiliated Hospital of the University of Sassari in Italy.