Computers, Materials & Continua CMC, vol.60, no.2, pp.511-525, 2019

Rigid Medical Image Registration Using Learning-Based Interest Points and Features

Maoyang Zou[1, 2], Jinrong Hu[2], Huan Zhang[2], Xi Wu[2], Jia He[2], Zhijie Xu[3] and Yong Zhong[1, *]

Abstract: For image-guided radiation therapy, radiosurgery, minimally invasive surgery, endoscopy and interventional radiology, one of the important techniques is medical image registration. In our study, we propose a learning-based approach named "FIP-CNNF" for rigid registration of medical image. Firstly, the pixel-level interest points are computed by the full convolution network (FCN) with self-supervise. Secondly, feature detection, descriptor and matching are trained by convolution neural network (CNN). Thirdly, random sample consensus (Ransac) is used to filter outliers, and the transformation parameters are found with the most inliers by iteratively fitting transforms. In addition, we propose "TrFIP-CNNF" which uses transfer learning and fine-tuning to boost performance of FIP-CNNF. The experiment is done with the dataset of nasopharyngeal carcinoma which is collected from West China Hospital. For the CT-CT and MR-MR image registration, TrFIP-CNNF performs better than scale invariant feature transform (SIFT) and FIP-CNNF slightly. For the CT-MR image registration, the precision, recall and target registration error (TRE) of the TrFIP-CNNF are much better than those of SIFT and FIP-CNNF, and even several times better than those of SIFT. The promising results are achieved by TrFIP-CNNF especially in the multimodal medical image registration, which demonstrates that a feasible approach can be built to improve image registration by using FCN interest points and CNN features.

.

Keywords: Medical image registration, CNN feature, interest point, deep learning.

1 Introduction

The purpose of image registration is to establish the corresponding relationship between two or more images, and the images are brought into the same coordinate system through transformation. For image-guided radiation therapy, radiosurgery, minimally invasive surgery, endoscopy and interventional radiology, one of the important techniques is image registration.

For image registration, intensity-based registration and features-based registration are two

[1] Chengdu Institute of Computer Application, University of Chinese Academy of Sciences, Chengdu, China.

[2] Chengdu University of Information Technology, Chengdu, China.

[3] School of Computing and Engineering, University of Huddersfield, UK.

[*] Corresponding Author: Yong Zhong. Email: 13981928503@139.com; zhongyong@casit.com.cn.

recognized approaches. The intensity-based image registration approach directly establishes the similarity measure function based on intensity information, and finally registers the images by using the corresponding transformation in the case of maximum similarity. There are classic algorithms of this approach such as cross-correlation, mutual information, sequence similarity detection algorithm and so on. In general, it can be used for rigid and non-rigid registration. Its registration precision is high correspondingly, but the speed is slow due to high computational complexity, and it is also troubled by the monotone texture. For the feature-based image registration approach, the images are registered by using the representative feature of the image. The classical feature-based image registration most commonly uses the feature of SIFT [Lowe (2004)] + Ransac filter, and the second commonly uses the speeded up robust features (SURF) [Bay, Tuytelaars and Gool (2006)] + Ransac filter. The matching pair coordinates are obtained by these approaches, so the image transformation parameters can be calculated. Compared with the intensity-based image registration approach, its computation cost is relatively low because it does not consider all the image regions, and it has stronger anti-interference ability and higher robustness to noise and deformation, but the precision of registration is generally lower. Overall, feature-based image registration approach is currently a hot research topic because of its good cost performance.

In recent years, the deep neural network which simulates human brain has achieved great success in image recognition [He, Zhang, Ren et al. (2015)], speech recognition [Hinton, Deng, Yu et al. (2012)], natural language [Abdel-Hamid, Mohamed, Jiang et al. (2014)], computer vision and so on [Meng, Rice, Wang et al. (2018)], and has become one of the hot research topics. In the task of computer vision classification [Krizhevsky, Sutskever and Hinton (2012)], segmentation [Long, Shelhamer and Darrell (2015)], target detection [Ren, He, Girshick et al. (2015)], the deep neural network, especially the Convolution Neural Network (CNN), performs well.

For medical image registration, features-based approaches are developed by deep neural network. Since Chen et al. [Chen, Wu and Liao (2016)] first register spinal ultrasound and CT images using CNN, the researchers have achieved some results with deep learning approaches in the registration of chest CT images [sokooti, Vos, Berendsen et al. (2017)], brain CT and MR images [Cheng, Zhang and Zheng (2018); Simonovsky, Gutieerrez-Becker, Mateus et al. (2016); Wu, Kim and Wang (2013); Cao, Yang and Zhang (2017)], 2D X-ray and 3D CT image [Miao, Wang and Liao(2016)], and so on. But overall, there are only a few researches on medical image registration using learning-based approach. Shan et al. [Shan, Guo, yan et al. (2018)] stated: "for learning-based approaches: (1) informative feature representations are difficult to obtain directly from learning and optimizing morphing or similarity function; (2) unlike image classification and segmentation, registration labels are difficult to collect. These two reasons limit the development of learning-based registration algorithms."

In this study, we propose a learning-based approach named "FIP-CNNF" to register medical images with deep-learning network. Firstly, FCN is used to detect the interest points in CT and MR images of nasopharyngeal carcinoma, which are collected from the patients in West China Hospital (This dataset is named "NPC"). Secondly, Matchnet network is used for feature detection, descriptor and matching. Thirdly, Ransac is used to

filter outliers, and then the CT-CT, MR-MR, CT-MR images are registered by iteratively fitting transforms to the data. In addition, transfer learning is adopted on FIP-CNNF (named "TrFIP-CNNF"). Specifically, the Matchnet network is pre-trained with UBC dataset to initialize network parameters, and then trained with NPC dataset. The experiment show that the registration results of TrFIP-CNNF are better than those of FIP-CNNF.

The contribution of this work is that:

- Two key steps of classic features-based registration algorithm have been improved by learning-based approach. A multi-scale, multihomography approach boosts pixel-level interest point detection with self-supervise, and Matchnet network using transfer learning contributes to feature detection, descriptor and matching.
- For CT-MR registration, the precision, recall, and TRE of TrFIP-CNNF are much better than those of SIFT. The result of experiment demonstrates that a feasible approach is built to improve multimodal medical image registration.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 mainly introduces the methodology. Section 4 describes the transfer learning. Section 5 presents the experimental setup and experimental results. Section 6 is the conclusion for this paper.

2 Related work

The feature-based image registration approach focuses on the features of the image. Therefore, it is the key to how to extract features with good invariance. SIFT is the most popular algorithm for feature detection and matching at present. The interest points found by SIFT in different spaces are very prominent, such as corner points, edge points, etc. The features of SIFT are invariance in rotation, illumination, affine and scale.

SURF is the most famous variant of SIFT, Bay et al. [Bay, Tuytelaars and Gool (2006)] proposed: "SURF approximates or even outperforms previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster."

The performance comparison of SIFT and SURF is given in Juan et al. [Juan and Gwun (2009)]. "SIFT is slow and not good at illumination changes, while it is invariant to rotation, scale changes and affine transformations. SURF is fast and has good performance as the same as SIFT, but it is not stable to rotation and illumination changes."

There are many other variants of SIFT algorithm, such as, Chen et al. [Chen and Shang (2016)] propose "an improved sift algorithm on characteristic statistical distributions and consistency constraint."

Although SIFT is widely used, it also has some shortcomings. For example, the SIFT requires that the image has enough texture when it constructs 128-dimensional vectors for interest points, otherwise the 128-dimensional vector constructed is not so distinguished that it is easy to cause mismatch.

CNN can also be used for feature extraction, feature description and matching. Given the image patches, the CNN usually employs the FC or pooled intermediate CNN features. The paper [Fischer and Dosovitskiy (2014)] "compares features from various layers of

convolutional neural nets to standard SIFT descriptors", "Surprisingly, convolutional neural networks clearly outperform SIFT on descriptor matching". Other approaches using CNN features include [Reddy and Babu(2015); Xie, Hong and Zhang (2015); Yang, Dan and Yang (2018)].

Here we specifically discuss Siamese network [Bromley, Guyon, LeCun et al. (1994)], which was first introduced in 1994 for signature verification. On the basis of Siamese network, combined with the spatial pyramid pool [He, Zhang, Ren et al. (2015)] (the network structure can generate a fixed-length representation regardless of image size/scale), Zagoruyko et al. [Zagoruyko and Komodakis (2015)] proposed a network structure of 2-channel + Central-surround two-stream + SPP to improve the precision of image registration. Han et al. [Han, Leung, Jia et al. (2015)] proposed "Matchnet" which is an improved Siamese network. By using fewer descriptors, Matchnet obtained better results for patch-based matching than those of SIFT and Siamese.

3 Methodology

This section focuses on the Methodology of FIP-CNNF. FIP-CNNF has three modules: (1) Interest point detection, (2) Feature detection, description, matching and (3) Transformation modelestimation, which will be described in detail as following.

3.1 Interest points detection

Inspired by Detone et al. [Detone, Malisiewicz and Rabinovich (2017)], we detect interest points in two steps. The first step is to build a simple geometric shapes dataset with no ambiguity in the interest point locations, which consists of rendered triangles, quadrilaterals, lines, cubes, checkerboards, and stars with ground truth corner locations. And then the FCN named "Base Detector" is trained with this dataset. The second step is finding interest points using Homographic Adaptation, and the process is shown in Fig. 1 [Detone, Malisiewicz and Rabinovich (2017)].

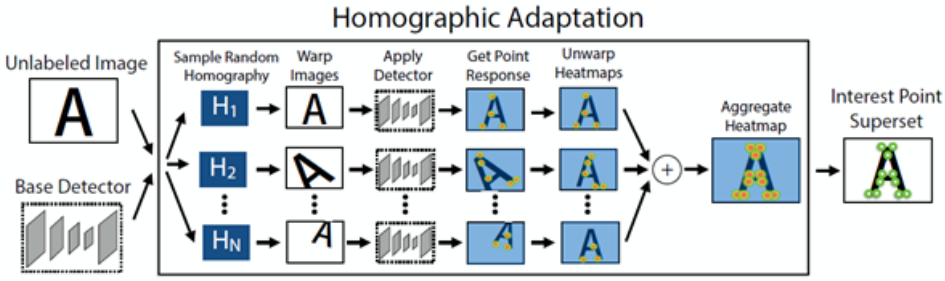

Figure 1: Homographic adaptation [Detone, Malisiewicz and Rabinovich (2017)]

To find more potential interest point locations on a diverse set of image textures and patterns, Homographic Adaptation applies random homographies to warp copies of the input image, so Base Detector is helped to see the scene from many different viewpoints and scales. After Base Detector detects the transformed image separately, the results is

combined to get the interest point of the image. The interest points from our experimental medical image are shown in Fig. 2 (the red interest points are obtained by SIFT and green interest points are obtained by homographic adaptation).

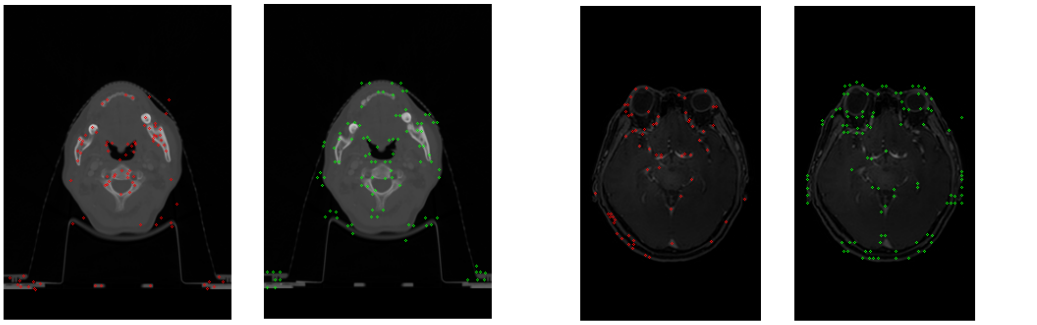

Figure 2: Interest points of CT and MR images

The ingenious design of this approach is that it can detect interest points with self-supervision, and it can boost interest point detection repeatability.

3.2 Feature detection, descriptor and matching

Siamese network can learn a similarity metric and match the samples of the unknown class with this similarity metric. For images that have detected interest points, feature detection, descriptor and matching can be carried out with a Siamese network. In our experiment, the deep learning network is called "Matchnet" which is a kind of improved Siamese network. The network structure is shown in Fig. 3 and the network parameters are shown in Tab. 1.

Figure 3: Network structure

Table 1: Network parameters

Name	Type	Output Dim	Patch Size	Stride
Conv1	Convolution	64*64*24	7*7	1
Pool1	Max Pooling	32*32*24	3*3	2
Conv2	Convolution	32*32*64	5*5	1
Pool2	Max Pooling	16*16*64	3*3	2
Conv3	Convolution	16*16*96	3*3	1
Conv4	Convolution	16*16*96	3*3	1
Conv5	Convolution	16*16*64	3*3	1
Pool5	Max Pooling	8*8*64	3*3	2
FC1 Full	Convolution	1024	-	-
FC2 Full	Convolution	516	-	-
FC3 Full	Convolution	2	-	-

The first layer of network is the preprocessing layer. "For each pixel in the input grayscale patch we normalize its intensity value x (in [0,255]) to (x-128)/160" [Han, Leung, Jia et al. (2015)]. For following convolution layers, Rectfied Linear Units (ReLU) is used as non-linearity. For the last layer, Softmax is used as activation function. The loss function of Matchnet is cross-entropy error, whose formula is as follow:

$$\mathrm{E} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (1)$$

Here training dataset has n patch pairs, y_i is the 0 or 1 label for input pair x_i, 0 indicates mismatch, 1 vice versa. $\hat{y}_i$ and 1- $\hat{y}_i$ are the Softmax activations computed on the values of v_0 (x_i) and v_1 (x_i) which are the two nodes in FC3, formula is as follow: $\hat{y}_i = \frac{e^{v_1(x_i)}}{e^{v_1(x_i)}+e^{v_0(x_i)}}$ (2)

$\hat{y}_i$ and 1- $\hat{y}_i$ are regarded as the possibility of two patches matching or mismatch respectively.

Formally, given a set S_1 of interest point descriptors in the fixed image, and a set S_2 of interest point descriptors in the moving image. For an interest point x in a fixed image, y_i is a corresponding point in a moving image, m is a measure of the similarity between the two points. The outputs of Matchnet network is a value between 0 and 1, and 1 indicates full match. To prevent matching when interest points are locally similar, which often occurs in medical images, we want to find the match between x and y_i is particularly distinctive. In particular, when we find the maximum m(x, y_1) and second largest m(x, y_2), the matching score is defined as:

$$\mathrm{h}(\mathrm{x}, s_2) = \frac{1-m(x,y_1)}{1-m(x,y_2)} \quad (3)$$

If h(x,S_2) is smaller, x is much closer to y_1 than any other member of S_2. Thus, we say that x matches y_1 if h(x,S_2) is below threshold η. In addition, it is considered that the

interest point x of the fixed image does not exist in the moving image if h (x,S_2) is higher than the threshold η .

We need to consider what the threshold is. When the threshold η is low, the real correspondence can be recognized less. After considering the effect on precision and recall under the η of 0.6,0.8,1.0 respectively, we define $\eta = 0.8$ in our experiment.

3.3 Transformation model estimation

The outliers of interest points are rejected by Ransac algorithm, and the transformation parameters are found with the most inliers by iteratively fitting transforms. The fixed image is transformed to the same coordinate system with the moving image. The coordinate points after image transformation are not necessarily integers, but we can solve this problem with interpolation.

4 Transfer learning

Greenspan et al. [Greenspan, Ginneken and Summers (2016)] have pointed out: "the lack of publicly available ground-truth data, and the difficulty in collecting such data per medical task, both cost-wise as well as time-wise, is a prohibitively limiting factor in the medical domain." Transfer learning and fine-tuning are used to solve the problem of insufficient training samples. Matchnet is pre-trained with UBC dataset which consists of corresponding patches sampled from 3D reconstructions of the Statue of Liberty (New York), Notre Dame (Paris) and Half Dome (Yosemite), and then the weights of the trained Matchnet are used as an initialization of a new same Matchnet, finally NPC dataset is used to fine-tune the learnable parameters of pre-trained Matchnet. According to the introduction of Zou et al. [Zou and Zhong (2018)]: "If half of last layers undergoes fine-tuning, compared with entire network involves in fine-tuning, the almost same accuracy can be achieved, but the convergence is more rapid", so half the last layers undergoes fine-tuning in our experiment.

5 Experiment

5.1 NPC dataset and data preprocessing

This study has been conducted using CT and MR images of 99 nasopharyngeal carcinoma patients(age range: 21-76 years; mean age $\pm$ standard deviation: 50.3 years $\pm$ 11.2 years) who underwent chemo radiotherapy or radiotherapy in West China Hospital, and the radiology department of West China Hospital agrees that this dataset is used and the experimental results can be published. There are 99 CT images and 99 MR images in NPC dataset, all of which are coded in DICOM format. The CT images are obtained by a Siemens SOMATOM Definition AS+ system, with a voxel size ranges from 0.88 mm*0.88 mm*3.0 mm to 0.97 mm*0.97 mm*3.0 mm. The MR images are obtained by a Philips Achieva 3T scanner. In this study, T1-weighted images are used, which have a high in-slice resolution of 0.61 mm*0.61 mm and a slice spacing of 0.8 mm.

The images are preprocessed as follows:

- Unifying the axis direction of MRI and CT data.

- Removing the invalid background area from CT and MR images.
- Unifying the images to have a voxel size of 1 mm*1 mm*1 mm.
- Because they are not consistent for the imaging ranges of MRI and CT, we only kept the range from eyebrow and chin when we slice the images.
- We randomly selected 15 pairs of MR and CT slices for each patient, and registered them as ground truth using the Elastix toolbox.

We augment the dataset by rotating and scaling.

- Rotation: rotating the slice by a degree from -15 to 15 with a step of 5.
- Scale: scaling the slice with a factor in [0.8, 1.2] with a step of 0.1.

We use the approach introduced in section 3.1 to detect the interest points, and then centring in the interest points, image patches of size 64*64 is extracted. If the patch pair is generated from the same or two corresponding slices and the absolute distance between their corresponding interest points is less than 50 mm, this patch pair receives a positive label; Otherwise, a negative label is obtained.

5.2 Experimental setup

The CT and MR images of 60 patients are used for training and validation, and 39 cases for testing. More than 2 million pairs of patch are produced in the way described in Section 5.1. From training and validation dataset, 500000 patch pairs are randomly selected as training data, 200000 patch pairs are used as validation data. From testing dataset, 300000 patch pairs are selected for testing. The ratio between positive and negative samples is 1: 1, and the proportion of MR-MR, CT-CT, CT-MR pairs is 1:1:2.

5.3 Results of experiment

The ground truth displacement at each voxel of test pairs is obtained by Elestix toolbox, so we can independently verify each matched interest Point, and then we can calculate the precision of the features extracted by SIFT, FIP-CNNF and TrFIP-CNNF respectively. True positive is matched interest point in the fixed image for a true correspondence exists, and false positives are interest point which is assigned an incorrect match.

For CT-CT image registration, the precision and recall of SIFT, FIP-CNNF and TrFIP-CNNF are shown as Fig. 4 and Fig. 5.

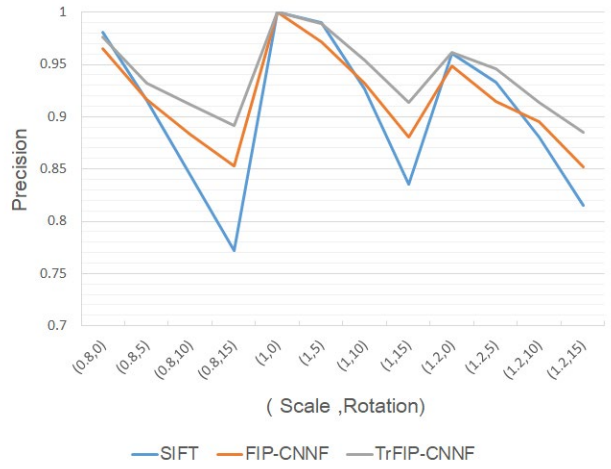

Figure 4: CT-CT Precision

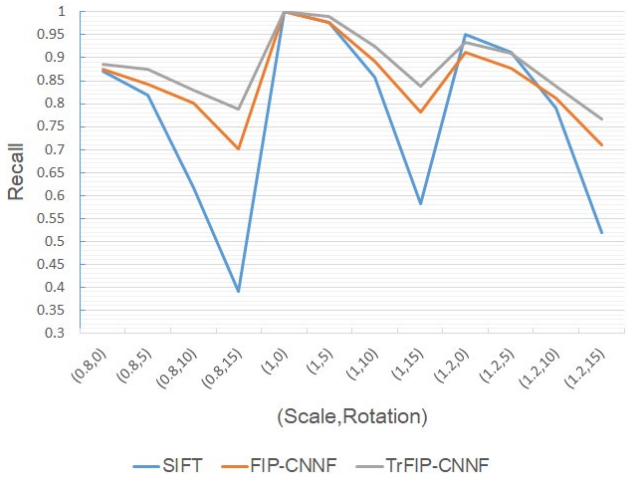

Figure 5: CT-CT Recall

X-coordinate (Scale, Rotation) represents the degree of the scale and rotation respectively. The experimental results show that TrFIP-CNNF outperforms SIFT and FIP-CNNF. For SIFT and FIP-CNNF, the mean value of the precision is little difference, and the recall of FIP-CNNF is better than that of SIFT.

For MR-MR image registration, the precision and recall of SIFT, FIP-CNNF and TrFIP-CNNF are shown as Fig. 6 and Fig. 7.

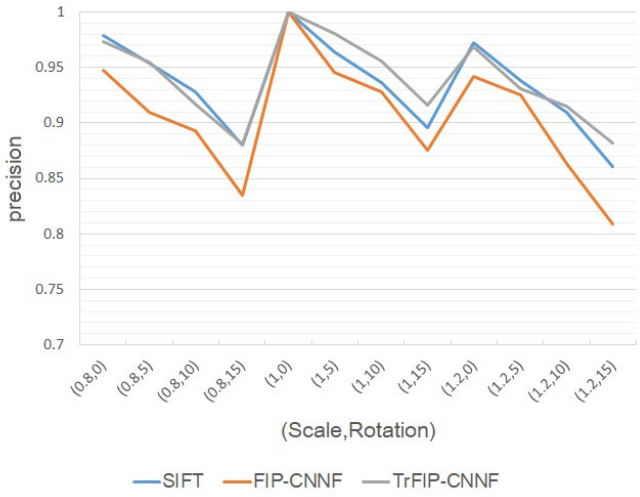

Figure 6: MR-MR Precision

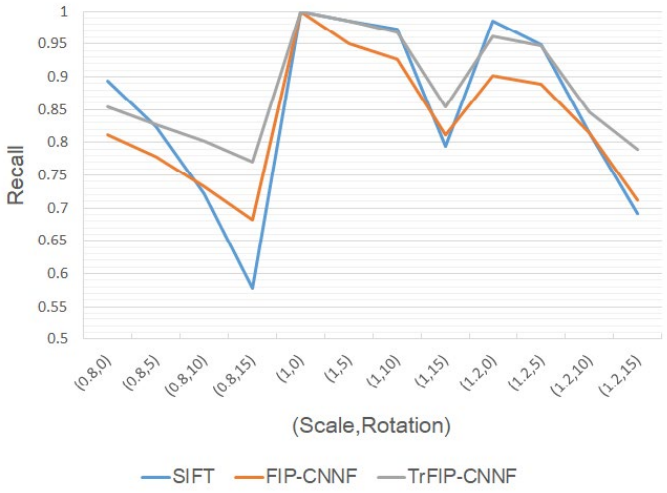

Figure 7: MR-MR Recall

The experimental results show that TrFIP-CNNF and SIFT perform well. In most cases, the precision and recall of TrFIP-CNNF are relatively higher when the rotation is greater than 5°, on the contrary, the precision and recall of SIFT algorithm are relatively higher when the rotation is less than 5°. Overall, the precision and recall of FIP-CNNF is the lowest.

For CT-MR image registration, the precision and recall of SIFT, FIP-CNNF and TrFIP-CNNF are shown as Fig. 8 and Fig. 9.

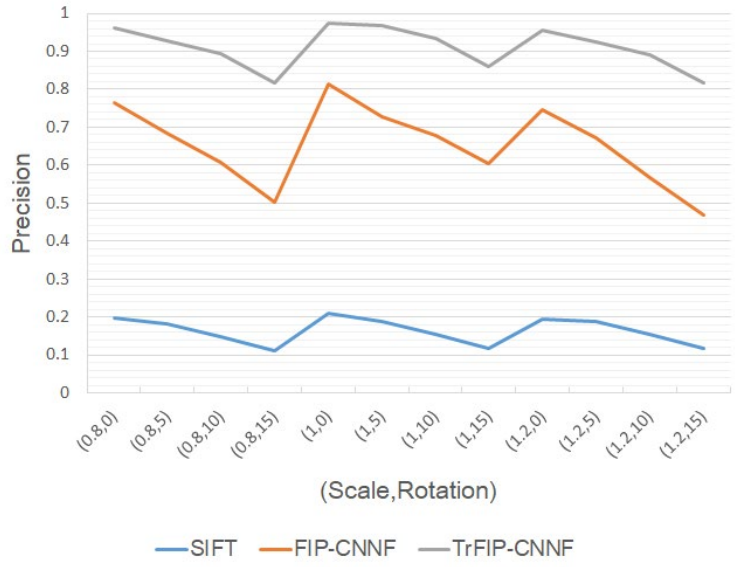

Figure 8: CT-MR Precision

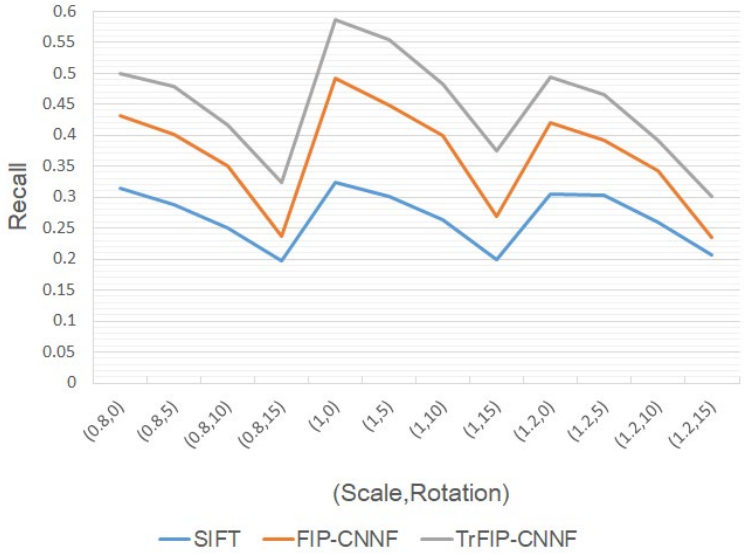

Figure 9: CT-MR Recall

For multimodal image registration, the deep learning approach has obvious advantages, so that FIP-CNNF and TrFIP-CNNF outperform SIFT in every task.

To further verify the results in Fig. 8 and Fig. 9, the target registration error (TRE) is calculated for measuring registration accuracy. TRE is defined as root-mean-square on these distance errors over all interest point pairs for one sample. TRE of multimode image registration are shown in Tab. 2. The first line (Scale, Rotation) represents the degree of the scale and rotation.

Table 2: TRE of CT-MR registration

	(0.8,0)	(0.8,5)	(0.8,10)	(0.8,15)	(1,0)	(1,5)	(1,10)	(1,15)	(1.2,0)	(1.2,5)	(1.2,10)	(1.2,15)
SIFT	1.256	2.048	3.881	5.337	1.245	1.421	2.954	4.951	1.981	2.042	3.053	5.132
FIP-CNNF	0.335	0.597	1.047	2.232	0.291	0.485	0.794	1.531	0.419	0.626	1.037	2.175
TrFIP-CNNF	0.015	0.016	0.031	0.087	0.010	0.012	0.028	0.053	0.011	0.011	0.021	0.082

It provides a visual comparison of a random pair of CT-MR slices registration between SIFT, FIP-CNNF and TrFIP-CNNF in Fig. 10.

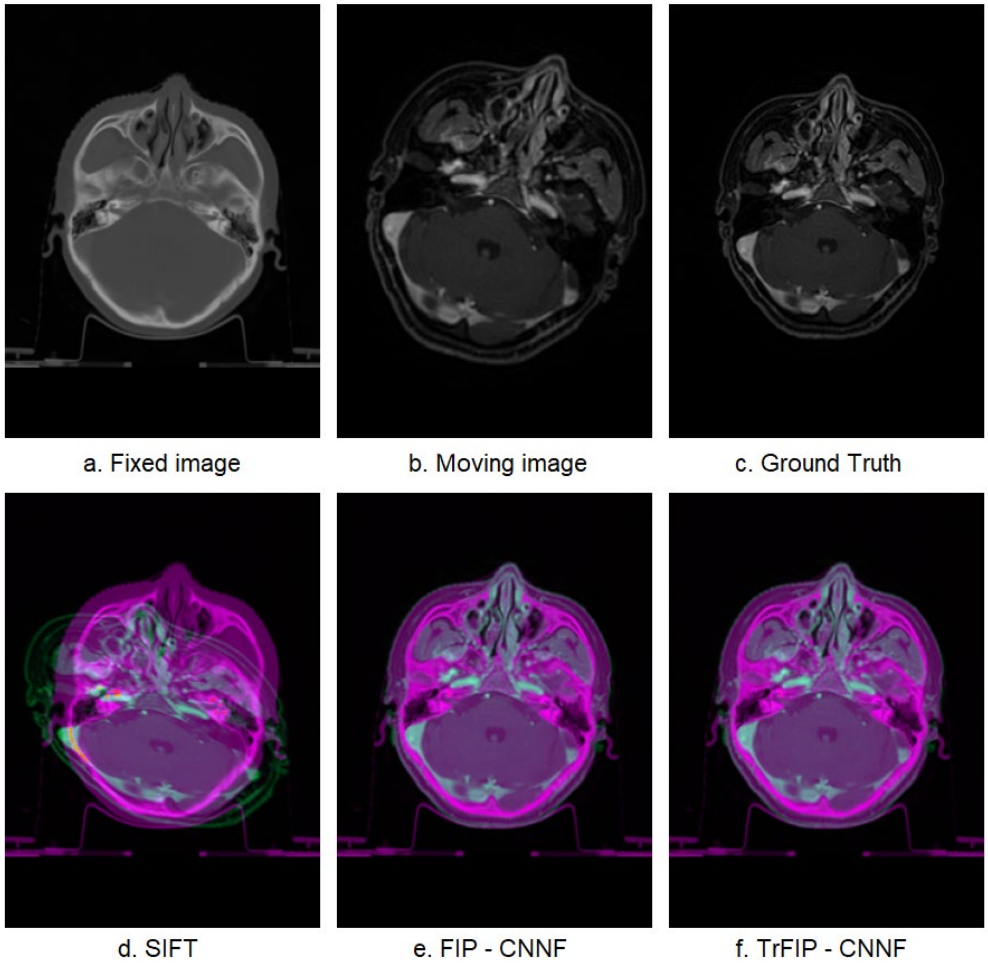

Figure 10: Color overlap registration results of SIFT, FIP-CNNF, and TrFIP-CNNF

6 Conclusion

In our study, the CT and MR images of nasopharyngeal carcinoma are registered by deep learning network. In particular, interest points are detected by FCN, and feature detection, descriptor and matching are trained by CNN. Experimental results show that this approach builds a general approach to improve medical image registration. Especially for the CT-MR image registration, FIP-CNNF outperforms SIFT in every task due to the superiority of the high level feature learned by CNN. TrFIP-CNNF outperforms FIP-CNNF due to the knowledge transferred by rich natural images, which indicates that transfer learning is feasible for medical image and fine-tuning has a positive impact.

Acknowledgement: We thank Xiaodong Yang for assistance in experiment. This work is supported by National Natural Science Foundation of China (Grant No. 61806029), Science and Technology Department of Sichuan Province (Grant No. 2017JY0011), and Education Department of Sichuan Province (Grant No. 17QNJJ0004).

References

Abdel-Hamid, O.; Mohamed, A. R.; Jiang, H.; Deng, L.; Penn, G. et al. (2014): Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio Speech Language Processing*, vol. 22, no. 10, pp. 1533-1545.

Bay, H.; Tuytelaars, T.; Gool, L. V. (2006): SURF: speeded up robust features. *European Conference on Computer Vision*, vol. 110, no. 3, pp. 404-417

Bromley, J.; Guyon, I.; LeCun, Y.; Sackinger, E.; Shah, R. (1993): Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence,* vol.7, no. 4, pp. 669-688.

Cao, X.; Yang, J.; Zhang, J.; Nie, D.; Kim, M.; Wang, Q. Shen, D. (2017): Deformable image registration based on similarity-steered CNN regression. *International Conference on Medical Image Computing Computer-Assisted Intervention*, vol. 10433, no.1, pp. 300-308.

Chen, F.; Wu, D.; Liao, H. (2016): Registration of CT and ultrasound images of the spine with neural network and orientation code mutual information. *International Conference on Medical Imaging and Virtual Reality*, vol. 9805, no. 1, pp. 292-301.

Chen, Y.; Shang, L. (2016): Improved sift image registration algorithm on characteristic statistical distributions and consistency constraint. *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 2, pp. 900-911.

Cheng, X.; Zhang, L.; Zheng, Y. (2018): Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 6, no. 3, pp. 248-252.

Detone, D.; Malisiewicz, T.; Rabinovich, A. (2018): Superpoint: self-supervised interest point detection and description. https://arxiv.org/abs/1712.07629.

Fischer, P.; Dosovitskiy, A.; Brox, T. (2014): Descriptor matching with convolutional neural networks: a comparison to sift. https://arxiv.org/abs/1405.5769.

Greenspan, H.; Ginneken, B. V.; Summers, R. M. (2016): Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153-1159.

Han, X.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg. A. C. (2015): Matchnet: unifying feature and metric learning for patch-based matching. *Computer Vision and Pattern Recognition*, pp. 3279-3286.

He, K.; Zhang, X.; Ren, S.; Sun. J. (2015): Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.

He, K.; Zhang, X.; Ren, S.; Sun. J. (2015): Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916.

Hinton, G.; Deng, L.; Yu, D.; Dahl, G. E.; Mohamed, A. et al. (2012): Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97.

Juan, L.; Gwun, O. (2009): A comparison of sift, PCA-SIFT and SURF. *International Journal of Image Processing*, vol. 3, no. 4, pp. 143-152.

Krizhevsky, A.; Sutskever, I.; Hinton. G. E. (2012): Imagenet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*, pp. 1097-1105.

Long, J.; Shelhamer, E.; Darrell, T. (2015): Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.

Lowe, D. G. (2004): Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110.

Meng, R.; Rice, S. G.; Wang, J.; Sun, X. (2018): A fusion steganographic algorithm based on faster R-CNN. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1-16.

Miao, S.; Wang, Z. J.; Liao. R. (2016): A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352-1363.

Mopuri, Reddy. K.; Babu, Venkatesh. R. (2015): Object level deep feature pooling for compact image representation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 62-70.

Ren, S.; He, K.; Girshick, R.; Sun. J. (2015): Faster R-CNN: towards real-time object detection with region proposal networks. *International Conference on Neural Information Processing Systems*, vol. 1, no. 1, pp. 91-99.

Shan, S.; Guo, X.; Yan, W.; Chang, E. I.; Fan, Y. et al. (2018): Unsupervised end-to-end learning for deformable medical image registration. https://arxiv.org/pdf/1711.08608.pdf.

Simonovsky, M.; Gutierrez-Becker, B.; Mateus, D.; Navab, N.; Komodakis, N. (2016): A deep metric for multimodal registration. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 9902, no. 1, pp. 10-18.

Sokooti, H.; de Vos, B.; Berendsen, F.; Lelieveldt, B. P.; Isgum, I. et al. (2017): Nonrigid image registration using multi-scale 3D convolutional neural networks.

International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 10433, no. 1, pp. 232-239.

Wu, G.; Kim, M.; Wang, Q. (2013): Unsupervised deep feature learning for deformable registration of MR brain images. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8150, no. 1, pp. 649-656.

Xie, L.; Hong, R.; Zhang, B.; Tian, Q. (2015): Image classification and retrieval are one. *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 3-10.

Yang, Z.; Dan, T.; Yang, Y. (2018): Multi-temporal remote sensing image registration using deep convolutional features. *IEEE Access*, vol. 6, no. 1, pp. 38544-38555.

Zagoruyko, S.; Komodakis, N. (2015): Learning to compare image patches via convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4353-4361.

Zou, M.; Zhong. Y. (2018): Transfer learning for classification of optical satellite image. *Sensing and Imaging*, vol. 19, no. 1, pp. 6.