Remote sensing image registration based on full convolution neural network and k-nearest neighbor ratio algorithm

View the article online for updates and enhancements.

Remote sensing image registration based on full convolution neural network and k-nearest neighbor ratio algorithm

Dongzhen Wang[1], Ying Chen[1] and Jipeng Li[1]

[1] Computer Science and Information Engineering, Engineering Shanghai Institute of Technology, Shanghai, China

1623527780@qq.com

Abstract. Aiming at the problem of low accuracy of remote sensing image registration caused by the negative effects of noise and imaging in some traditional algorithms, an effective remote sensing image registration method based on depth feature from coarse to fine is proposed. In the coarse registration stage, the full convolution neural network is used to extract the features of the input image, then the nearest neighbor distance ratio algorithm is used to coarse match the features and finally an approximate transformation matrix is obtained. In the stage of fine registration, firstly, the image features are extracted by the improved convolutional neural network based on shortcut connection, then the affine transform coefficients are obtained by the combination of approximate transform matrix and k-nearest neighbor ratio algorithm. Finally, the image to be registered can be transformed according to the coefficients to achieve the purpose of registration. Experimental results show that, compared with the comparison method, the proposed method can increase the correct matching correspondence, so as to improve the accuracy of registration.

1. Introduction

For different remote sensing images, for most feature-based sift , the traditional registration algorithm is difficult to ensure the accuracy and robustness of one-step registration. Ma w et al. [1] proposed in 2019 that registration should be divided into two parts to solve this problem. The first step is rough registration and the approximate spatial relationship is obtained by convolution neural network. The second step is fine registration and the matching strategy considering spatial relationship is applied to the method based on local features.

For the feature extraction stage of rough registration, if only the depth feature extracted from the last layer of convolutional neural network is used to complete the image registration, the final output result will be affected. In 2015, abdulkadir a et al. [2] proposed the U-net network. The first half of the U-net network is for feature extraction and the second half is for upsampling. U-net adopts completely different multi-scale feature fusion methods, that is, the upsampling part will fuse the output of the feature extraction part. In rough registration, the nearest neighbor distance ratio (NNDR) is used for feature matching. Firstly, the ratio of the nearest neighbor to the next neighbor is calculated, then a reasonable threshold is set. When the distance is greater than the threshold, it is judged as mismatching, otherwise it is judged as matching.

For the fine registration stage, in the feature extraction stage, the general convolution neural network improved by shortcut link [3] is used for feature extraction and then combined with the previous approximate transformation matrix, the reference point set will get an approximate matching point in the set of points to be matched. In the feature matching stage, k nearest neighbors (KNN)

algorithm is used to calculate the nearest k points between the point set to be matched and the reference point set and the closest point between the approximate matching point and the k points is the best matching point.

2. This paper introduces the framework of registration system

In order to achieve image registration, this paper proposes a remote sensing image registration model based on depth feature from coarse to fine. In this paper, the network model is divided into the following two parts: (1) rough registration. Firstly, the image pairs to be registered are input into the full convolution neural network for feature extraction; then the NNDR is used for initial matching; finally, the approximate transformation matrix is obtained according to the matching relationship. (2) Fine registration, using the general convolution neural network to extract the features of the input image; then the reference point set combined with the approximate transformation matrix to get the approximate location of the best matching point in the point set to be matched; finally, the KNN ratio nearest algorithm is used to find the best matching point among the first k neighbors closest to the reference point in the point set to be matched Match. The accurate transformation matrix is calculated according to the final correspondence and the image to be registered is transformed according to the transformation matrix to complete the registration. The flow chart of the related registration system is shown in Figure 1.
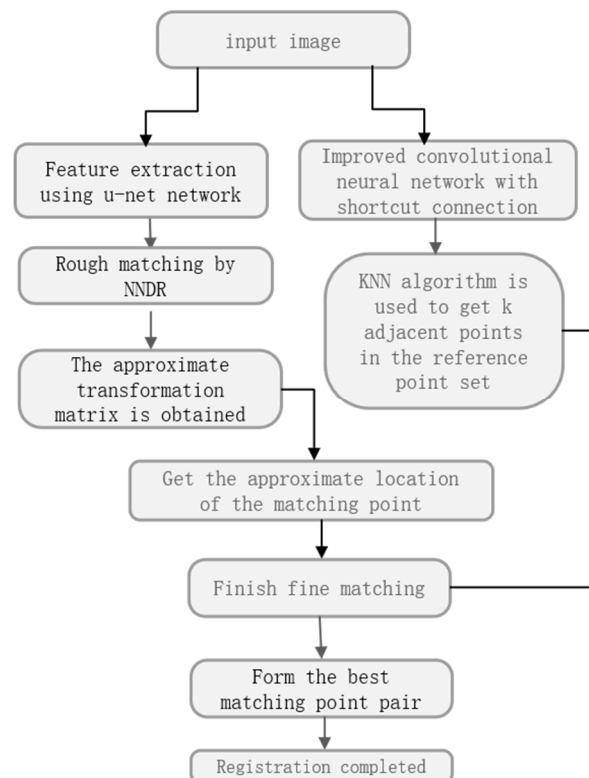

Figure 1. Flow chart of registration system in this paper

3. Coarse registration

3.1. Full convolution neural network

In this paper, VGG16 is proposed as the main framework of U-net network to improve. The improved network model can be divided into two parts: compressed path and extended path. The left half of the improved U-net network architecture is the compressed path, which is used to downsampling the input image several times in VGG16 network and the right half of the image is the extended path, which is used to upsampling the obtained feature image several times. Part of the layers between the two paths

are fused, so that the shallow features extracted from the left network are fused with the deep features extracted from the right network, which increases the richness of the features extracted from the network and the final image registration effect is better.

VGG16 is a compressed path in the improved U-net network model. The network contains five convolution $blocki(i = 1,2,3,4,5)$, each convolution block contains several convolution layers and one pooling layer [4]. The convolution kernel size of each convolution layer is 3×3, the step size is 1×1 and the number of convolution kernels in each convolution block is 64,128,256,512,512. The step size of pooling layer is 2×2 and the pooling mode is maximum pooling. The calculation formula of feature map scale is as follows:

$$output = \frac{input + 2 * padding - kernel\_size}{stride} + 1 \qquad (1)$$

Among them, $output$: the size of the output feature map after convolution; $input$: the size of the input image; $padding$: the number of circles around the input image; $padding = 1$ means adding a circle around the input image; $kerne\_size$: the size of the convolution kernel; $stride$: the step size of each convolution movement when the convolution kernel convolutes the input image.

The extended path in the improved U-net network model also contains five $up\ samplingi(i = 1,2,3,4,5)$, each upsampling convolution block contains several convolution layers and one up sampling layer. Compared with the traditional U-net network with symmetrical network structure, the improved U-net network in this paper directly connects to the extended path at the end of the compressed path, omits part of the convolution transition operation, reduces the loss of features with the deepening of the network and also speeds up the operation speed of the network. The calculation formula of feature map scale is as follows:

$$output' = stride * (input - 1) + kernel\_size - 2 * padding \qquad (2)$$

Where, $output'$: the size of the output feature map after upsampling.

3.2. Nearest neighbor distance ratio

In coarse registration, the matching stage can be calculated by the nearest neighbor distance ratio, which is usually obtained by euclidean distance. This method has been applied to many remote sensing image feature matching:

$$NN_1/NN_2 < d_{ratio} \qquad (3)$$

Among them, we define NN_1 is the nearest value between the points in the reference image and *the image to be registered*; NN_2 is the value of the second nearest distance between the points in the reference image mapped to the image to be registered; d_{ratio} represents the threshold value, which can be set manually. When the ratio of NN_1, NN_2 is less than the threshold, it indicates that the current matching point meets the matching requirements. On the contrary, it indicates that the point is an outer point and does not meet the requirements of matching.

4. Fine registration

4.1. Improved general convolutional neural network with shortcut connection

The traditional feedforward neural network takes the output of the n layer as the input of the $n + 1$ layer directly. Suppose the network contains n layers, each layer contains a nonlinear transformation $H_n = (*)$, where n is the n layer in the network. $H_n = (*)$ represents various network functions, such as convolution or pooling, normalization. Mark the output of the n layer as X_n. Then the transformation of traditional network can be expressed as:

$$X_n = H_n(X_{n-1}) \qquad (4)$$

The shortcut link network adds a residual link to the feed-forward neural network and the new output X'_n can be expressed as:

$$X'_n = H_n(X_{n-1}) + X_{n-1} \qquad (5)$$

The general convolutional neural network used in the fine registration method in this paper also contains five convolution blocks, each of which is connected by two or four convolution layers and

one pooling layer in sequence. Some of the parameters are as follows: the convolution kernel size of each convolution layer is 3x3, the step size is 1x1 and the number of convolution kernels of each convolution block is 64,128,256,512,512. The step size of pooling layer is 2x2 and the pooling mode also adopts maximum pooling.

4.2. K nearest neighbor distance ratio nearest

In the feature matching stage of fine registration, we use KNN (k nearest neighbors) algorithm and approximate transformation matrix, that is, knnric (k nearest neighbor ratio is closest) to further adjust the matching results of coarse registration. KNN (k nearest neighbor) is the k-nearest neighbor algorithm. The so-called k-nearest neighbor algorithm, as the name suggests, is to express its meaning according to its nearest K neighbors, which means that each sample can be represented by the category of the nearest K neighbors. Its principle is that when predicting which category a new sample point belongs to, the sample point will be judged according to which category it belongs to among the nearest K neighbors Which category does it belong to.

In KNN, the distance between objects is calculated as the dissimilarity index between objects to avoid the matching problem between objects:

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

Where ρ is the euclidean distance between two points; x_1, y_1 is the coordinate value of the sample point; x_2, y_2 is the coordinate value of the adjacent point.

K nearest neighbor distance ratio nearest (knnric) is a partial improvement or special case on the basis of KNN algorithm. When predicting which category a new sample point belongs to, this paper selects k neighboring points for K classification, that is, by calculating the distance between the comparison sample point and the k points to get the matching points. The formula is defined as follows:

$$\begin{cases} d_{min} = min[\rho_1, \rho_2, \cdots, \rho_K] \\ \rho_i = \sqrt{(x_i - x_1)^2 - (y_i - y_1)^2} \end{cases} \tag{7}$$

Where, d_{min} is the minimum distance between the sample point and K adjacent points $\rho_i (i = 1,2,\cdots,K)$ denotes the distance between the sample point and K adjacent points; x_1, y_1 represents the coordinate value of the sample point, here refers to the approximate transformation point generated by the reference point and the approximate transformation matrix; x_i, y_i refers to the coordinate value of the nearest point to the reference point.

In this paper, the reference points generated by convolution neural network are combined with the approximate transformation matrix. The specific steps are shown in the below. The reference image generating point o and obtains the approximate matching point p in the image to be registered according to the approximate transformation matrix obtained by rough registration. The procedure of k-nearest distance ratio nearest matching algorithm is to first calculate the k points which are closest to the reference point o in the image to be registered (the best matching point is usually one of these points), then calculate the values of the k points which are close to the approximate matching point p, and finally compare the k values obtained and find out the point with the smallest value, which is the best matching point of this algorithm point.

5. Experimental analysis

5.1. Data set and parameter setting

Firstly, the full convolution neural network model based on VGG as the main framework and the improved ordinary convolution neural network model are built, then the model is trained with the constructed data set. In this paper, 59 groups of multi temporal remote sensing image pairs are used to construct the dataset, and Z matching "seed" image blocks are selected on 59 pairs of images. Each of these "seed" image block pairs is declared to represent its own class. To extend these classes, k = 250 random transforms are applied to each "seed" image block pair. Each transform is a combination of

several random basic transforms, including rotation, translation, scaling and brightness. Therefore, the custom data set constructed in this paper has Z classes, each class contains 2K samples, and is randomly divided into training set and test data set according to the ratio of 4:1.

5.2. The experimental results are presented in this paper

In the registration experiment, this paper selects three types of test images of city, desert and lake in aid remote sensing image data set to test the registration results of this method, the results are shown in the figure below. Among them, the resolutions of the three image pairs are 678x672, 1352x1382 and 752x726 respectively.

（a）Reference image

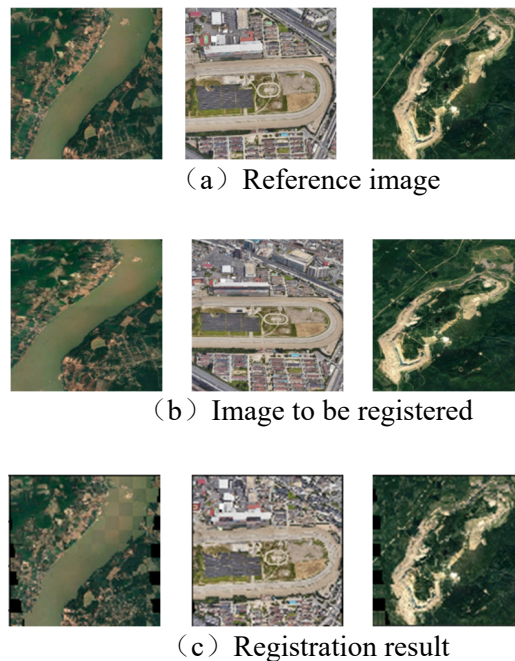（b）Image to be registered

（c）Registration result

Figure 2. Experimental test of image registration results

5.3. Contrast experiment

In order to objectively compare and analyze, this paper compares two other traditional registration algorithms and a deep learning image registration method and carries out three groups of experiments under different test image pairs. Compared with other methods, the checkerboard method is used to display the final results, as shown in Figure 3 below.

The three groups of test image pairs are all transformed in the shooting view angle, which makes the image pairs have certain geomorphic differences. The traditional algorithms such as SIFT algorithm and reference [5] may have more exterior points or insufficient feature points when extracting features from image pairs with certain geomorphic differences, so the registration effect is obviously inferior to the deep learning algorithm. In reference [6], only VGG single channel network is used for image registration. In this paper, the VGG network is improved and extended path is added to form a full convolutional neural network, so that the network can fuse multi-scale features to complete rough image registration. Then, the improved ordinary convolutional neural network with shortcut connection is used to increase the feature reuse rate and k-nearest distance ratio of the image the value nearest matching algorithm adjusts the result of coarse registration. It can be seen from the comparison that the registration effect of this method is better than that of other methods in local areas.
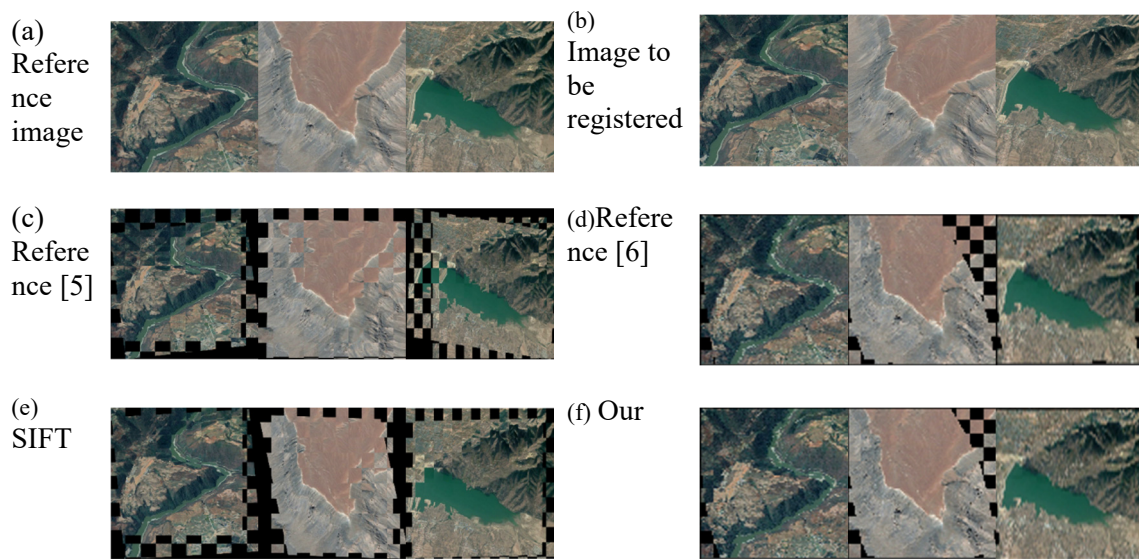
Figure 3. The experimental results were compared

6. Conclusion

In view of the fact that the accuracy and robustness of partial registration algorithm can not be guaranteed by one-step registration, this paper proposes a method of coarse registration first and then fine registration. In the coarse registration stage, aiming at the problem of insufficient features detected by the general neural network, this paper proposes to use the full convolution neural network to extract the features of the input image and then use the nearest neighbor distance ratio for feature matching to generate the corresponding approximate transformation matrix to complete the coarse registration. In the stage of fine registration, the improved convolution neural network with shortcut connection is used for feature extraction and then the k-nearest distance ratio algorithm is used to further optimize and adjust the results. Experiments show that the registration results of the proposed method are better than other comparison methods.

References

[1] Ma W , Zhang J , Wu Y , et al. A Novel Two-Step Registration Method for Remote Sensing Images Based on Deep and Local Features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019:1-10.

[2] iek, zgün, Abdulkadir A , Lienkamp S S , et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[J]. 2016.

[3] Huang G , Liu Z , Laurens V D M , et al. Densely Connected Convolutional Networks[J]. 2016.

[4] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[5] Ma W , Wen Z , Wu Y , et al. Remote Sensing Image Registration With Modified SIFT and Enhanced Feature Matching[J]. IEEE Geoscience and Remote Sensing Letters, 2016, PP(99):1-5.

[6] Z. Yang, T. Dan and Y. Yang, "Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features," in IEEE Access, vol. 6, pp. 38544-38555, 2018.