International Journal of Computer Applications Technology and Research
Volume 12–Issue 03, 06-08, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1203.1002

UVA Image Registration Model Based on VGG and Multi-Branch Attention

Jieyuan Luo
School of Communication Engineering
Chengdu University of Information Technology
Chengdu, China

Penjing Dong
School of Communication Engineering
Chengdu University of Information Technology
Chengdu, China

Qinglin Huang
School of Communication Engineering
Chengdu University of Information Technology
Chengdu, China

Abstract:
For UVA images with different resolutions and large areas of weak texture, image feature extraction is insufficient and mis-matching is increased during image registration. To solve these problems, an unsupervised registration model based on VGG feature extraction and multi-branch attention is proposed. First of all, two feature extraction networks with shared weight parameters are used to extract the low and high level fusion features of the moving image and the reference image. The convolution neural network is used to extract the high-dimensional feature map of the image, and the key points are selected according to the conditions that meet both the channel maximum and the local maximum, and the corresponding 512-dimensional descriptor is extracted on the feature map, In the matching stage, add multi-branch attention based on residual block to filter out the wrong features. The algorithm is tested with multiple groups of images and compared with several image matching algorithms. The results show that the algorithm can extract the scale-invariant similar features of images, and has strong adaptability and robustness.

Keywords: deep learning; image matching; convolution neural network; unsupervised learning; multi-branch attention

1. INTRODUCTION

At present, image registration is one of the essential key technologies in the process of image mosaic. Image registration methods include gray-level based registration methods and feature-based registration methods [1]. Among them, gray-level based methods complete image registration through gray-level value calculation. This method is simple and intuitive, but the calculation amount is large and sensitive to the gray-level value of the image, and the illumination change of the image Scale change and rotation change will cause large matching error; The feature-based registration method obtains the registration results by extracting and matching the common features between images to calculate the transformation parameters. This method has good robustness and high efficiency. The illumination and inclination of different UAV images often differ greatly, so it is more appropriate to use feature-based registration method.

Feature-based image registration methods can be subdivided into traditional methods and learning-based methods The typical traditional method is the SIFT (Scale Invariant Feature Transform) algorithm proposed by D.G. Lower et al. [2] The algorithm performs registration by extracting scale, scale and rotation invariance features, which has stable performance but high complexity and is sensitive to mismatched data Although a series of optimization algorithms [3] have been generated for this algorithm, they all have certain scenario constraints and computational efficiency is not high.

In recent years, deep learning methods have shown excellent performance in the field of image [4]. Many researchers have used deep learning methods such as convolutional neural networks (CNN) to solve image registration problems [5] In order to solve the problem of lack of tag images in deep learning, some scholars explored unsupervised learning registration method VoxelMorph [6] method has achieved good results on brain data sets; VTN (Volume Tweening Network) adopts integrated affine transformation module and network block cascade mode, and has achieved success in medical image registration with large deformation; Literature [7] used unsupervised learning of photometric loss to estimate homography; Literature [8] added mask structure to learn the depth information of the image after feature extraction, so as to make more accurate homography estimation, and so on . In general, image registration based on depth learning is becoming the mainstream However, due to the large resolution and large area of weak texture areas of aerial images taken by UAVs, it is easy to cause feature mis-matching, thus reducing the registration accuracy. Therefore, at present, there are few studies on applying depth learning model to such image registration.

2. Design of feature extraction module

The feature extraction module design, as the first step of the registration model design in this paper, is mainly to use the high performance of deep learning to extract the advanced feature information of the image pair to be registered, so as to achieve robust and efficient feature alignment. In view of the excellent performance of VGG-16 network on ImageNet, the front part of VGG-16 network structure is used to extract features. However, VGG structure has no branch structure, In the shallow network part, the low-level contour features of the image are extracted, while in the deep network part, the high-level detail information is filtered. Simply stacking the network can not combine the low-level and high-level features. Therefore, using a simple VGG network structure can not effectively extract the features that are conducive to image registration. The ResNet structure can apply the output of the previous layer to the next layer, The low-level contour features and high-level semantic features can be fused, but the ResNet series network is deep and complex, and the image registration task requires a relatively simple model to ensure the operational efficiency. Therefore, this paper combines the

ResNet idea with the VGG network structure. It can not only screen out the low and high level fusion features required for registration, but also ensure that the network structure is relatively simple. The specific network structure is shown in Figure1.

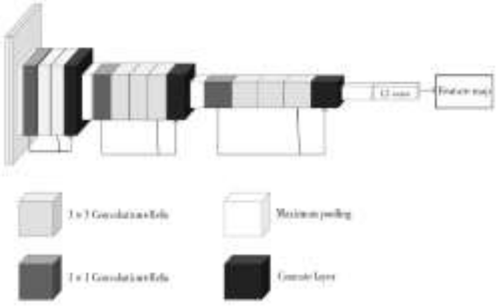

Figure. 1 Feature extraction network structure diagram

The image with the input resolution (H×W) size is first passed through two convolution kernel sizes of 3 and channel number 64 to obtain Conv1; The pooling operation of Conv1 makes the image resolution 1/2 of the original image to reduce the dimension. For pooling results (Pool1), use (1×1) convolution to increase the number of channels to 128 to obtain r1; Conv2 is obtained by convolution of r1 with two convolution kernel size of 3, step size of 1 and channel number of 128; R1 and Conv2 are added in the channel dimension, so that the output of the previous layer is applied to the next layer to achieve the effect of feature fusion. The subsequent network structure is so on, the number of channels of convolution is 256,512, the resolution is 1/4, 1/8 of the original image, each convolutional layer is followed by a modified linear unit (Relu), and (1×1) convolution is carried out after each pooling, the result is applied to the next layer, the network is cut to (pool4), and finally the feature map is L2 normalized.

3. Feature matching module design based on Multi-branch attention

The feature matching layer is used to calculate all similarity pairs between the local descriptors of the moving image feature map f_M and the reference image feature map f_F. Preliminary feature matching can be achieved by using the correlation layer [6], but due to the existence of a large area of weak texture areas (such as water, sky, etc.) in the UAV image, it is easy to cause wrong feature matching in the feature matching stage, so a multi-branch attention module is added to filter the wrong feature matching to enhance the robustness of the model outliers.

The initial matching partial correlation layer is input to two feature maps f_M and $f_{F,}$ and output a three-dimensional correlation diagram $C_{FM} \in R^{H\times W\times(H\times W)}$, and define each element on the position (i, j, k) as a pair of corresponding positions The scalar product of a descriptor, the mathematical description of which is:

$$C_{FM}(i,j,k)=f_M(i,j)^{T}f_F(i_k,j_k)$$

i∈{1 , …, W}, j∈{1 , …, H}, k∈{1 , …, W × H}. (i , j) and (i_k , j_k) refer to dense features at H×W;A single feature location in the diagram; k＝H（j_k-1）+ i_k, (ik,jk), that is, each of length W ×H correlation Vector; C_{FM} (i,j,k) stands for f_M The neutral coordinate is the local of (i,j).Descriptor with $f_{F.}$ The degree of similarity between local descriptors in. The design idea of multi-branch attention module to filter out false matching is to take the correlation graph C_{FM} as input and output the weight matrix W with the same resolution as C_{FM}, in which the corresponding position weight value of the correct match is larger and the weight of the corresponding position of the wrong match is small. After this, the original is related The C_{FM} figure is weighted by the weight matrix W, and the value at the correct match is increased and the value at the false match is decreased. On this basis, an attention network composed of two parallel branches is designed, and two weight plots W1 and W2 are generated respectively. In Multi-branch module, each branch consists of two parts, encoding and decoding, using the residual element as the basic unit, and the basic structure of the residual element is shown in Figure 2.

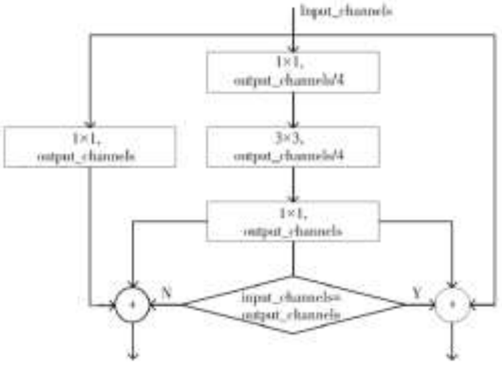

Figure.2 Residual unit structure information graph

4. Experimental results and analysis

4.1 Experimental parameter settings

The overall network is designed using the TensorFlow framework, and UAV-123[9] is used to form 2k registration image pairs, including buildings, roads, cars, sailboats and other categories. Divide all pairs of images to be registered into three parts, namely training set, verification set and test set, with a ratio of 0.75:0.05:0.2. With the help of NVIDIA TITAN X GPU server training network, the initial learning rate selected in training is 0.000 1, attenuation is 10% every 10 rounds, and the batch size is set to 4, for a total of 50 rounds of training. After several experiments, the weight λ of perceived loss in the loss function is finally set to 10, and the neural network is trained until convergence using Adam optimizer.

4.2 Mean Absolute Error MAE

Mean Absolute Error MAE (Mean Absolute Error) represents the average absolute error of a pixel position, which is a general form of error average. When doing model evaluation, it has better robustness to outliers. The smaller the value of MAE, the more similar the two images, that is, the better the registration effect x_i, y_i Indicates the registration result image and the reference image, respectively the pixel value at the i position; N represents the total number of pixels.

$$MAE=\frac{\sum_{i=1}^{N}|y_i-x_i|}{N}$$

4.3 Objective indicators evaluation analysis

Table 1. Statistics evaluation indicators different methods on the test set

method	MAE	Time(CPU)/s
SIFT	161.3254	1.09
ORB	196.4541	0.37
UBHE	178.5633	0.53
CAU-DHE	170.4625	1.99
R-VGG	**140.0882**	**0.88**

It can be seen from Table 1 that the proposed method achieves the best results in MAE indicators, followed by SIFT algorithm and lowest indicators of ORB algorithm, which is consistent with the results of subjective observation. In addition, the indicators of all methods in the table are at low values, mainly due to the large difference between the pairs of images to be registered, the overlapping range of the reference image and the moving image is small, and the registration result image has a large area of black edges. The overall evaluation index of the proposed method is high and the calculation time is short, which proves that the proposed method is non-existent Effectiveness on human-machine image registration tasks.

5. Conclusion

In this paper, an image registration model for UAV based on unsupervised learning is proposed. Firstly, making full use of the high performance of deep learning, the R-VGG feature extraction module is designed to screen out the low- and high-level fusion features with robust characteristics. Secondly, the feature matching module adds Multi-branch attention (MBA) constraints are introduced to filter out false matches, thereby improving registration accuracy. In addition, the composite loss function weighted by content loss and perceived loss is used to improve network performance, and the analysis of visual perception and objective indicators is verified Effectiveness and stability of the method in the field of UAV aerial image registration. In future work, the analysis of drone images will be studied The depth information is improved to make full use of the image information to improve the registration accuracy.

6. REFERENCES

[1] LONG Yongzhi. Research on infrared and visible image registration and fusion algorithms[D] Chengdu : University of Electronic Science and Technology of China,2020.

[2] Lowe D G . Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[3] Qy A , Dn A , Yj A , et al. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171:1-17.

[4] Bay H , Tuytelaars T , Gool L V . SURF: Speeded up robust features[C]// Proceedings of the 9th European conference on Computer Vision - Volume Part I. Springer-Verlag, 2006.Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[5] Ye F , Su Y , Hui X , et al. Remote Sensing Image Registration Using Convolutional Neural Network Features[J]. IEEE Geoscience & Remote Sensing Letters, 2018, PP(2):1-5.

[6] Rocco I , Sivic J . Convolutional neural network architecture for geometric matching[J]. IEEE Computer Society, 2017.

[7] Nguyen T , Chen S W , Shivakumar S S , et al. Unsupervised Deep Homography: A Fast and Robust Homography Estimation Model[C]// International Conference on Robotics and Automation. IEEE, 2018.

[8] Zhang J , Wang C , Liu S , et al. Content-Aware Unsupervised Deep Homography Estimation:, 10.48550/arXiv.1909.05983[P]. 2019.

[9] Leibe B , Matas J , Sebe N , et al. [Lecture Notes in Computer Science] Computer Vision – ECCV 2016 Volume 9905 || A Benchmark and Simulator for UAV Tracking[J]. 2016, 10.1007/978-3-319-46448-0(Chapter 27):445-461.