



OPEN

Robust registration of SAR and optical images based on deep learning and improved Harris algorithm

Wannan Zhang

Traditional algorithms can achieve good results when registering homologous images, but it cannot reach satisfying results for registration between synthetic aperture radar (SAR) and optical images. The difficulty is that the image texture information and structures of different modalities is very different which leads to poor registration results. To solve this problem, we present a robust matching framework for registration between SAR and optical images. First, a novel deep learning network is utilized to generate high quality pseudo-optical images from SAR images. Next, feature points are detected and extracted using the multi-scale Harris algorithm. Then the feature points are constructed through the gradient position orientation histogram method. Finally, the actual position of the feature points will be reconstructed through a feedback mechanism for matching. Experimental results demonstrate its superior matching performance with respect to the state-of-the-art methods.

Image registration is a process of aligning two images of the same scene so that corresponding pixels can get the same coordinates¹⁻⁵. This research has been widely used in many practical applications, especially in the field of remote sensing such as change detection, loss assessment, image fusion, and post-disaster rescue. In recent years, with the increasing of high-resolution SAR image data, the registration of SAR and optical images has gradually become a popular topic⁶.

Traditional image registration methods^{7,8} generally include three categories: (1) Feature-based image registration methods, including SIFT-based and SURF-based registration algorithms, etc.; (2) Region-based image registration methods, including MI-based and CCRE-based registration algorithms, etc.; (3) Local structural similarity based methods, including HOPC-based and HIOHC-based algorithms. Although these traditional algorithms can achieve good results when registering homologous images, it cannot achieve satisfying results for registration between SAR and optical images. This is because the image texture information and structures of different modalities are very different which results in poor registration results.

Recently, deep learning (Deep Learning, DL) has begun to emerge in various fields and convolutional neural network (CNN) has been widely used in the area of image processing for its outstanding performance⁹. The amount of CNN-based image processing methods has grown dramatically such as regional convolutional neural network features (Regions with CNN features, R-CNN), region-based fast convolutional neural network (Fast Region-based Convolutional Network, Fast R-CNN), and single-layer multi-frame detectors (Single Shot Multi Box Detector, SSD) and Deep Residual Network (Res Net), etc. These deep learning network models extract and combine different levels of image features¹⁰. One advantage of this mechanism is that it can realize self-learning of features. Therefore, we introduce the deep learning network for translating SAR images into pseudo-optical images first and then realize registration.

In this paper, we propose a robust matching framework for registration between SAR and optical images. First, a novel deep learning network is utilized to generate high quality pseudo-optical images from SAR images. Next, feature points are detected and extracted using the multi-scale Harris algorithm. Then the feature points are constructed through the GLOH method. Finally, the actual position of the feature points will be reconstructed through a feedback mechanism for matching.

School of Computer, Central South University, Changsha, China. email: 1508622762@qq.com

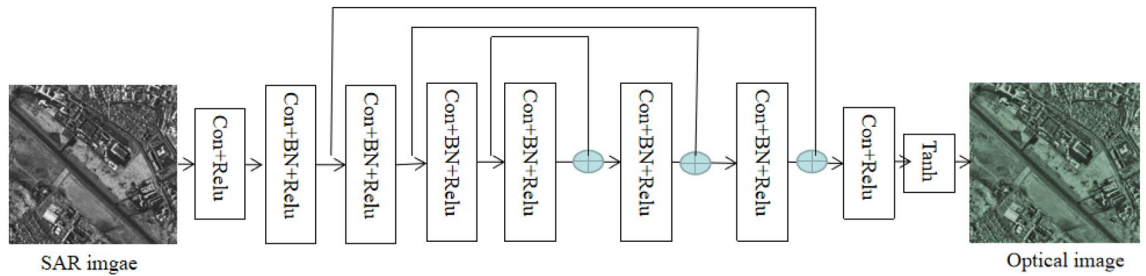


Figure 1. Proposed network for SAR to pseudo-optical image translation.

Methodology

Proposed network for SAR to optical image translation. In this section, we provide details of the proposed deep learning framework for generating pseudo-optical images from SAR images. The network consists of two main components: colorization network and generative adversarial learning. In the colorization network, we introduce an adversarial loss for better image colorization.

Deep learning-based image colorization has been studied over the last couple of years^{11,12}. Fully leverage the contextual information of an image is the key step during an image colorization neural network for color translation. Generally, an encoder-decoder architecture is added for extracting and utilizing the contextual information. The input image is encoded into a set of feature maps in the middle of the network. But this means that all information flows need pass through all the layers during such a network. Considering the image colorization problem, the sharing of low-level information between the input and output is important since the input and output should share the location of prominent edges. For the above reasons, we add skip connections which is following the general shape of an encoder-decoder CNN as shown in Fig. 1. The colorization sub-network forms a symmetric encoder-decoder with 8 convolution layers and 3 skip connections. For each convolution layer, the kernel size is 3 × 3.

As for the translation of SAR images, one important part is that the output image must be noise free and realistic¹³. One common loss function used in many image translation problems is the L₁ loss. Although the L₁ loss has been shown to be very effective for image de-noising problem, it will incentivize an average, grayish color if it is uncertain which of several plausible color values a pixel should take on. In particular, L₁ will be minimized by choosing the median of the conditional probability density function over possible colors. Thus, the L₁ loss alone is not suitable for image colorization. Recent studies have shown that the adversarial loss can become aware that gray looking outputs are unrealistic, and encourage matching the true color distribution. Considering the pros and cons of both losses, we combine the per-pixel L₁ loss and the adversarial loss together with appropriate weights to form our new refined loss function.

Perform gradient calculation and feature point extraction on the image. The image gradient must be calculated before the feature point extraction. The edge detection Sobel operator can quickly calculate the direction convolution kernel which is required for key point detection of the subsequent Harris algorithm¹⁴. First define two templates in the horizontal and vertical direction as:

$$f_H = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, f_V = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{1}$$

Use the two templates in Eq. (1) to convolve with the image gray value I(x, y) to get the gradient values in the horizontal and vertical directions. Taking into account the scale invariance, the scale parameter α_i is introduced, and f_H and f_V.

are regarded as the volume of two rectangular sub-windows and Gaussian kernel functions. The multi-scale Sobel operator used can be expressed as:

$$F_{H,\alpha_i} = G_{\alpha_i} * f_H, F_{V,\alpha_i} = G_{\alpha_i} * f_V \tag{2}$$

In the formula, F_H, α_i, F_V, and α_i are the gradients in the horizontal and vertical directions respectively, G_{α_i} is the Gaussian kernel function corresponding to α_i, and * represents the convolution operation. The scales in the optical image and the SAR image correspond to each other, satisfying;

$$\frac{\alpha_{i+1}}{\alpha_i} = k \tag{3}$$

Therefore, the gradient size and direction can be expressed as:

$$\begin{cases} F_{M,\alpha_i} = \sqrt{(F_{H,\alpha_i})^2 + (F_{V,\alpha_i})^2} \\ F_{O,\alpha_i} = \arctan(\frac{F_{V,\alpha_i}}{F_{H,\alpha_i}}) \end{cases} \tag{4}$$

In the formula, F_{M,α_i} is the gradient magnitude matrix of the image, and F_{O,α_i} is the gradient direction matrix.

When the original SIFT algorithm detects the key points of SAR images, the multiplicative speckle noise will have a serious impact on the second derivative used which results in that reliable key points cannot be detected¹⁵. Therefore, the key point detection method is improved during the SIFT algorithm. Experiments show that the multi-scale Harris detection method can detect key points with higher repeatability and stronger stability, which is better and much faster than the minimum nuclear similarity zone (SUSAN) isocenter detection. Based on the gradient calculation, multi-scale Harris function is used to construct the scale space. The candidate key points of each layer are extracted by calculating the local maximum value, and non-maximum value suppression is performed¹⁶. The multi-scale Harris function can be expressed as:

$$M(\alpha_i) = G_{\sqrt{2}\alpha_i} * \begin{bmatrix} (G_{H,\alpha_i})^2 & (G_{H,\alpha_i}) \cdot (G_{V,\alpha_i}) \\ (G_{V,\alpha_i}) \cdot (G_{H,\alpha_i}) & (G_{V,\alpha_i})^2 \end{bmatrix} \quad (5)$$

$$R(\alpha_i) = \det [M(\alpha_i)] - dtr[M(\alpha_i)]^2 \quad (6)$$

where: α_i is the scale of the image, G_H , α_i , G_V and α_i are the horizontal and vertical gradients on the scale α_i respectively, d is any parameter, \det is the value of the matrix determinant, tr is the trace of the matrix, and R is the scale space.

Construct descriptors and perform feature matching. After feature point detection, the GLOH¹⁷ method is used to establish the descriptor. This descriptor can improve the processing speed of the algorithm while retaining more structural information of the image. It solves the problem of inconsistencies in the main directions of heterogeneous images which is caused by the traditional descriptor creation method, and making the final registration result more stable. At the same time, the nearest neighbor distance ratio (NNDR)¹⁸ method is used to measure the similarity between descriptors and the FSC (Fast sample consensus) algorithm¹⁹ is used to delete the wrong matching point pairs.

Reconstruct feature points of the original image. Considering the problem of image de-redundancy will cause the lack of image pixels and the output image quality is changed when the de-redundant image is directly used for registration, we propose the feature point reconstruction method to make the final registration order and the target of the segment is the original input image²⁰. The core idea of feature point reconstruction is that the descriptor is used after de-redundancy to restore the coordinate information in the original image, then compare the deleted elements and coordinate information in set Ω and Ω' recorded during the de-redundancy process, and calculate the total number of rows and columns removed before the current coordinates. The coordinates of the corresponding points in the original image are the sum of horizontal and vertical coordinates of the feature points in the redundant image, and the number of rows and columns are removed. The process of feature point reconstruction algorithm:

1. Enter the description of the redundant image $P = \{p_1, p_2, \dots, p_x\}$, extract the descriptors of the visible light image and the SAR image;
2. Compare the coordinate information in Θ and Θ' with the row number and column number recorded in Ω and Ω' in turn. Take Θ and Ω as an example, the comparison method: arrange all the i_{nums} in Ω in ascending order, and use $p_{i,1}$ in Θ for interpolation sorting. The size of $p_{i,1}$ is the number of rows i_{row} that were removed before that point. The number of columns that were eliminated before the point i_{col} .
3. Repeat step 2) for other descriptors to obtain the coordinates in the original image. Taking the i th feature point as an example, the coordinates in the original image are:

$$(q_{i,1}, q_{i,2}) = (p_{i+1} + i_{\text{row}}, p_{i,2} + i_{\text{col}}) \quad (7)$$

4. Obtain the position information of feature points of the original image, and perform the parameter estimation of the affine transformation model based on these feature points, then the model is finally to complete the image registration correction.

Experimental results and analysis

To evaluate the performance of the proposed method, three pairs of SAR and optical images are experimented. The experiments are compiled with Python3.6, and the network is built through the deep learning framework of Pytorch1.3, and the corresponding CUDA10.0 and cudnn7.0 are configured for GPU acceleration. The test data consists of different characteristics including different resolutions, incidence angles, seasons etc. The dataset description is shown in Table 1. Experimental results are shown in Figs. 2, 3, 4 and Table 2.

To quantitatively evaluate the registration performances, we adopt the root-mean-square error (RMSE)²¹ between the corresponding matching keypoints, and it can be expressed as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (5)$$

where (x_i, y_i) and (x'_i, y'_i) are the coordinates of the i th matching keypoint pair; n means the total number of matching points. In addition, correct matching ratio (CMR) is another effective measure which is defined as:

Image No	Image source	Size/(pixel × pixel)	Spatial resolution/m	Date	Location
1	TerraSAR-X	580 × 520	2.5	07/2018	Urban area
	Google Earth	580 × 520	3	05/2017	
2	TerraSAR-X	650 × 500	3	12/2010	River area
	Google Earth	650 × 500	3	09/2012	
3	TerraSAR-X	550 × 460	2	10/2018	Suburb area
	Google Earth	550 × 460	3	04/2018	

Table 1. Detailed description of dataset.

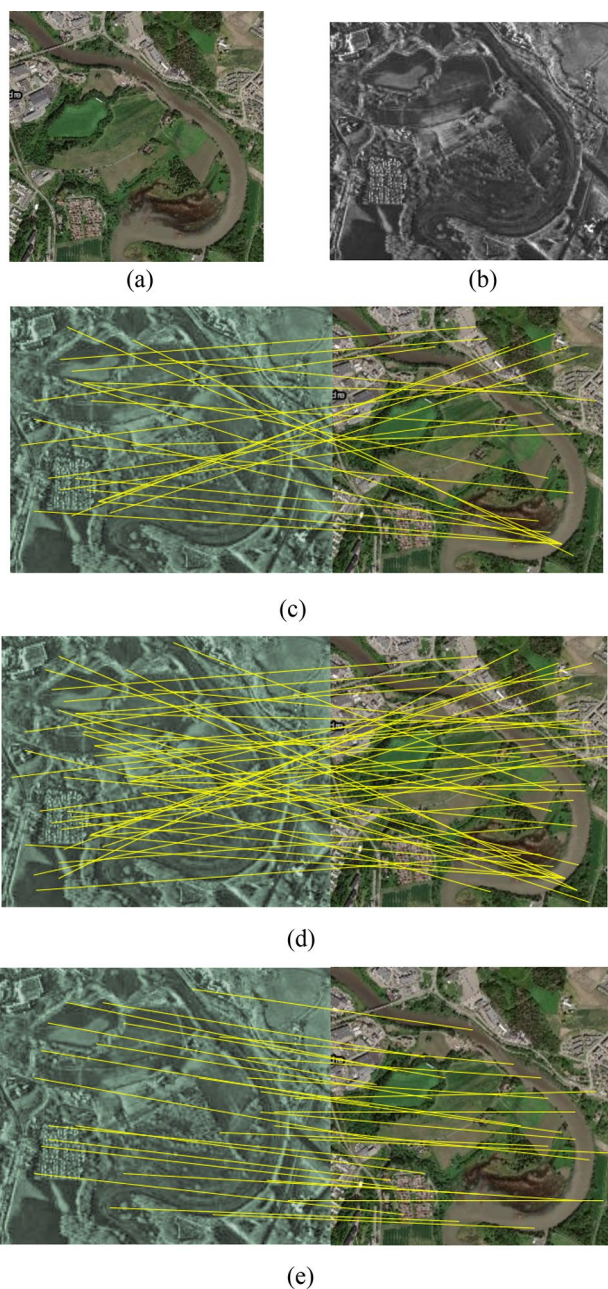


Figure 2. (a) Optical image; (b) SAR image; Matches found in pair 1 using (c) SIFT-M, (d) PSO-SIFT, and (e) the proposed method.

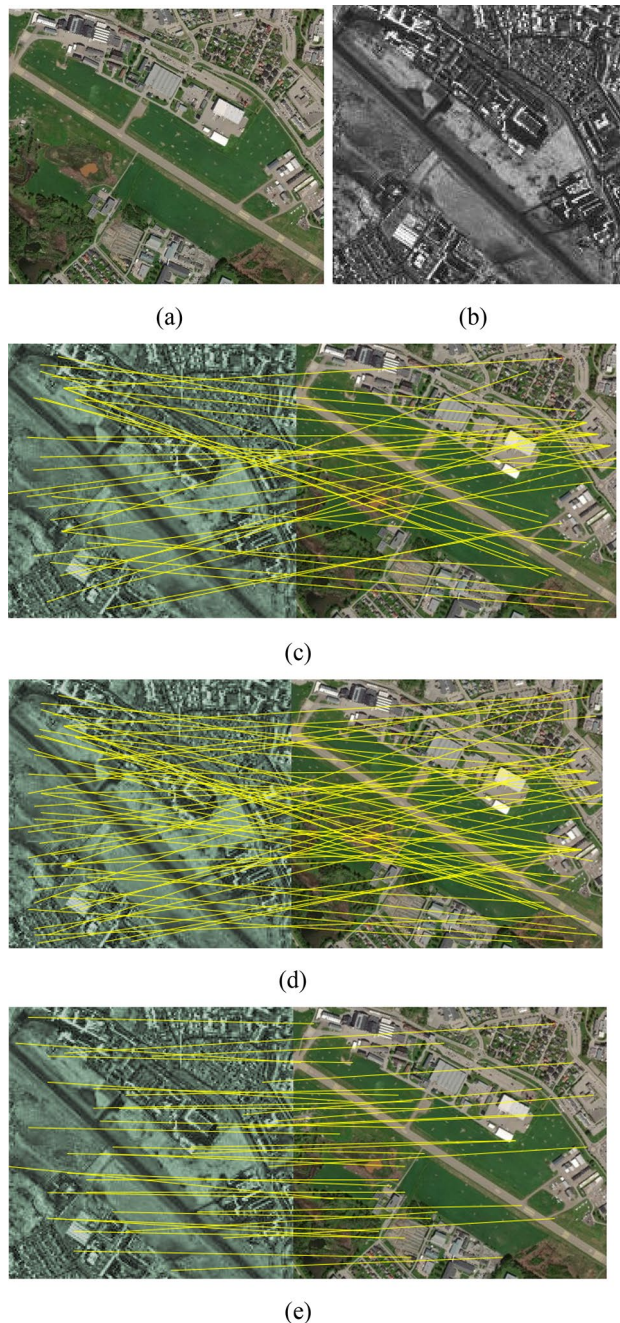


Figure 3. (a) Optical image; (b) SAR image; Matches found in pair 2 using (c) SIFT-M, (d) PSO-SIFT, and (e) the proposed method.

$$CMR = \frac{\text{correct Matches}}{\text{correspondences}} \quad (6)$$

“correspondences” is the number of matches after using PROSAC, “correctMatches” is the number of correct matches after removing false ones. The results of quantitative evaluation for each method are listed in Table 2.

It can be seen from Table 2 that the SIFT algorithm fails to match in heterogeneous image registration, and the correct matching rate obtained by the SIFT-M¹⁹ and PSO-SIFT²⁰ algorithms is relatively low, and the PSO-SIFT algorithm runs relatively fast. After a certain rule of de-redundancy of the image, the number of feature point pairs for registration can be greatly reduced. The original image reconstruction of the feature point pairs before the affine transformation model estimation can ensure the accuracy of heterogeneous image registration. Therefore, the proposed algorithm reduces greatly the running time as well as improves the efficiency of SAR and optical image registration.

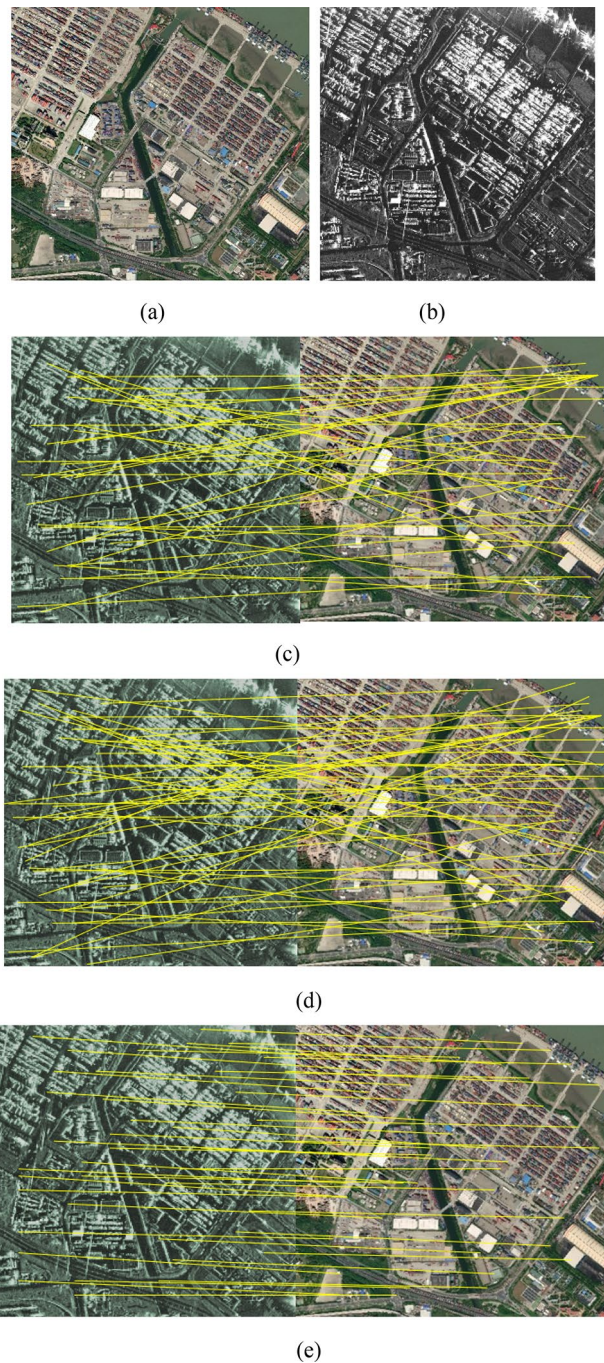


Figure 4. (a) Optical image; (b) SAR image; Matches found in pair 3 using (c) SIFT-M, (d) PSO-SIFT, and (e) the proposed method.

Conclusion

In this paper, we present a robust matching framework for registration between SAR and optical images. First, a novel deep learning network is utilized to generate high quality pseudo-optical images from SAR images. Next, feature points are detected and extracted using the multi-scale Harris algorithm. Then the feature points are constructed through the GLOH method. Finally, the actual position of the feature points will be reconstructed through a feedback mechanism for matching. Experimental results demonstrate its superior matching performance with respect to the state-of-the-art methods. Future work will mainly comprise a CNN-based framework for learning to identify corresponding patches in SAR and optical images in a fully automatic manner.

Image No	Method	CMR/%	RMSE/pixel	Running Time/s
1	SURF	22.12	3.8331	16.92
	SIFT-M	67.05	1.3782	51.27
	PSO-SIFT	74.96	1.0921	46.73
	Proposed	82.73	0.6014	39.28
2	SURF	25.79	3.4552	17.69
	SIFT-M	78.16	1.4701	53.62
	PSO-SIFT	76.28	1.5714	45.29
	Proposed	85.53	1.0182	39.54
3	SURF	20.56	3.9302	10.72
	SIFT-M	62.29	0.8751	32.49
	PSO-SIFT	53.63	1.0753	25.34
	Proposed	72.98	0.5912	20.18

Table 2. Quantitative comparison of the proposed method with other SIFT-based algorithms.

Received: 9 October 2021; Accepted: 31 March 2022

Published online: 07 April 2022

References

- Ye, Y. *et al.* Robust registration of multimodal remote sensing images based on structural similarity. *IEEE Trans. Geosci. Remote Sens.* **55**(5), 2941–2958 (2017).
- Pallotta, L. *et al.* Subpixel SAR image registration through parabolic interpolation of the 2-D cross correlation. *IEEE Trans. Geosci. Remote Sens.* **58**(6), 4132–4144 (2020).
- Sansosti, E. *et al.* Geometrical SAR image registration. *IEEE Trans. Geosci. Remote Sens.* **44**(10), 2861–2870 (2006).
- Pallotta, L., *et al.* SAR coregistration by robust selection of extended targets and iterative outlier cancellation. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5. <https://doi.org/10.1109/LGRS.2021.3132661> (2021).
- Merkle, N. *et al.* A new approach for optical and SAR satellite image registration. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2**, 119–126 (2015).
- Ye, Y. *et al.* Fast and robust matching for multimodal remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **57**(11), 9059–9070 (2019).
- Tuia, D., Marcos, D. & Camps-Valls, G. Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. *ISPRS J. Photogramm. Remote Sens.* **120**, 1–12 (2016).
- Zhang, Y. Understanding image fusion. *Photogram. Eng. Remote Sens.* **70**(6), 657–661 (2004).
- Mao, X., Shen, C. & Yang, Y.-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems* 2802–2810 (2016).
- Brunner, D., Lemoine, G. & Bruzzone, L. Earthquake damage assessment of buildings using VHR optical and SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **48**(5), 2403–2420 (2010).
- Zhang, R., Isola, P. & Efros, A. A. Colorful image colorization. In *European Conference on Computer Vision* 649–666. (Springer, 2016).
- Iizuka, S., Simo-Serra, E. & Ishikawa, H. Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans. Graph.* **35**(4), 110:1–110:11 (2016).
- Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. arXiv preprint [arXiv:1611.07004](https://arxiv.org/abs/1611.07004) (2016).
- Gonçalves, H., Gonçalves, J. A., Corte-Real, L. & Teodoro, A. C. CHAIR: Automatic image registration based on correlation and Hough transform. *Int. J. Remote Sens.* **33**(24), 7936–7968 (2012).
- Kingma, D. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- Bunting, P., Labrosse, F. & Lucas, R. A multi-resolution area-based technique for automatic multi-modal image registration. *Image Vis. Comput.* **28**(8), 1203–1219 (2010).
- Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) (2015).
- Uss, M. L., Vozel, B., Lukin, V. V. & Chehdi, K. Multimodal remote sensing image registration with accuracy estimation at local and global scales. *IEEE Trans. Geosci. Remote Sens.* **54**(11), 6587–6605 (2016).
- Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- Fan, B. *et al.* Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geosci. Remote Sens. Lett.* **10**(4), 657–661 (2013).
- Ma, W. P. *et al.* Remote sensing image registration with modified SIFT and enhanced feature matching. *IEEE Geosci. Remote Sens. Lett.* **14**(1), 3–7 (2017).

Author contributions

W.Z. wrote the whole manuscript.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022