ISPRS Journal of Photogrammetry and Remote Sensing 218 (2024) 187–203

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Joint block adjustment and variational optimization for global and local radiometric normalization toward multiple remote sensing image mosaicking

Dekun Lin [a], Huanfeng Shen [a,b,c,*], Xinghua Li [d], Chao Zeng [a], Tao Jiang [a], Yongming Ma [a], Mingjie Xu [a]

[a] School of Resource and Environmental Sciences, Wuhan University, PR China
[b] Key Laboratory of Geographic Information System of Ministry of Education, PR China
[c] Key Laboratory of Digital Mapping and Land Information Application of the Ministry of Natural Resources, PR China
[d] School of Remote Sensing and Information Engineering, Wuhan University, PR China

ARTICLE INFO

Keywords:
radiometric normalization (RN)
Global and local
Block adjustment
Variational optimization
Remote sensing image mosaicking

ABSTRACT

Multi-temporal optical remote sensing images acquired from cross-sensor platforms often show significant radiometric differences, posing challenges when mosaicking images. These challenges include inconsistent global radiometric tones, unsmooth local radiometric transitions, and visible seamlines. In this paper, to address these challenges, we propose a two-stage approach for global and local radiometric normalization (RN) using joint block adjustment and variational optimization. In the first stage, a block adjustment based global RN (BAGRN) model is established to simultaneously perform global RN on all the images, eliminating global radiometric differences and achieving overall radiometric tonal consistency. In the second stage, a variational optimization based local RN (VOLRN) model is introduced to address the remaining local radiometric differences after global RN. The VOLRN model applies local RN to all the image blocks within a unified energy function and imposes the ℓ_1 norm constraint on the data fidelity term, providing the model with a more flexible local RN capability to radiometrically normalize the intersection and transition areas of the images. Therefore, the local radiometric discontinuities and edge artifacts can be eliminated, resulting in natural and smooth local radiometric transitions. The experimental results obtained on five challenging datasets of cross-sensor and multi-temporal remote sensing images demonstrate that the proposed approach excels in both visual quality and quantitative metrics. The proposed approach effectively eliminates global and local radiometric differences, preserves image gradients well, and has high processing efficiency. As a result, it outperforms the state-of-the-art RN approaches.

1. Introduction

In recent years, mosaicked images covering a wide range of regions of interest, generated from multiple remote sensing images, have become increasingly important in various fields, such as geographic mapping, resource and environmental monitoring, and disaster monitoring (Li et al., 2020b; West et al., 2019; Sui et al., 2020). However, factors such as the solar incidence angle, atmospheric conditions, lighting conditions, and acquisition time often cause global and local radiometric differences between images during remote sensing imaging (Chen et al., 2005; Yu et al., 2017). These differences manifest as bright and dark patterns, discontinuous radiometry, and unnatural transitions in the images, making seamless mosaicked image generation challenging. Therefore, to eliminate these radiometric differences, radiometric normalization (RN) must be performed between the multiple images before stitching. In the literature, RN is also referred to as radiometric equalization, balancing, correction, and color correction, transfer, or stabilization.

In RN, it is generally assumed that the reflection conditions in the overlapping region do not change. RN involves establishing a mapping relationship based on the radiometric correspondence in the overlapping region and generalizing it to the entire image, to achieve RN of multiple images. RN methods for multiple images can be classified into three categories: global models, local models, and combined models (Li

---

* Corresponding author.
E-mail address: shenhf@whu.edu.cn (H. Shen).

https://doi.org/10.1016/j.isprsjprs.2024.08.016
Received 7 June 2024; Received in revised form 22 August 2024; Accepted 29 August 2024
0924-2716/© 2024 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

et al., 2019b).

The global model based approaches (Brown and Lowe, 2007; Xiong and Pulli, 2010; Cresson and Saint-Geours, 2015; Shen et al., 2016; Xia et al., 2017b) express the radiometric mapping relationship between images as a global linear or nonlinear model, solving the RN coefficients of each image by a path propagation based or global optimization based approach. The path propagation based approaches are frame-by-frame color transfer methods(Li et al., 2015; Reinhard et al., 2001; Xiang et al., 2009) that correct the target image based on a given reference image. Examples of the path propagation based approaches are the combination of the Voronoi diagram and Dijkstra shortest path method (Pan et al., 2010; Li et al., 2019a) and the minimum spanning tree method (Xia et al., 2017a). These approaches work well when the image content is simple and the differences in imaging conditions are small, but struggle with more complex scenarios and larger numbers of images as the cumulative errors become larger. To avoid the issues encountered with the path propagation based approaches, the global optimization based approaches incorporate radiometric correspondence into a unified optimization model to obtain a simultaneous solution. These approaches include the quadratic optimization method (Cresson and Saint-Geours, 2015), the block adjustment based method (Yu et al., 2017; Liu et al., 2020; Zhang et al., 2022), the multi-surface model combination method (Zhou, 2015), and the maximum *a posteriori* based method (Li et al., 2022). These global optimization based approaches can overcome the cumulative error and uneven error distribution of the path propagation based approaches, often achieving a better overall radiometric consistency. However, the global models can effectively eliminate the radiometric differences in images with simple scenes, low spatial resolution, and small imaging differences, but struggle with local radiometric differences in complex scenes, small overlapping areas, and large imaging differences, resulting in local feature discontinuities and edge artifacts.

In contrast to the global models, the local model based approaches adopt the idea of subregions or classification, establishing local RN models based on subregional or categorized information to obtain RN coefficients for each region or category, achieving fine local RN. For example, Tai et al. (2005) divided the image into multiple regions based on probabilistic segmentation, statistically characterized the radiometry of each region as a component of a Gaussian mixture model, and used an improved expectation–maximization method to obtain the radiometric mapping function. However, this approach can only achieve natural radiometric transitions in simple scenes, is very time-consuming, and requires a parameter to determine the desired regions. Xiang et al. (2009) optimized the expectation–maximization method based on Tai et al. (2005), supported the use of multiple reference images to improve the color transfer in complex scenes, and improved the computational efficiency. Li et al., 2015 used the local moment matching method for local RN and obtained good results, but this method can only perform RN between two images. Oliveira et al. (2015) used mean shift to segment the image into regions, followed by a region fusion algorithm to extract the connected regions. Local joint image histograms based on all the regions were then modeled as truncated Gaussians, and maximum likelihood estimation was used to solve the local color palette mapping functions. In addition, Zhang et al. (2022) introduced normalized difference vegetation index (NDVI) data into the model for local RN by different feature classes. In summary, the local models can usually eliminate the local radiometric differences and achieve natural and smooth radiometric transitions. However, these approaches are often insufficient to achieve overall tonal consistency across multiple images. In addition, if the segmentation or classification results do not match the local features of the images, new radiometric differences may be introduced, leading to obvious differences and transition discontinuities between different regions or classes.

The combined model based approaches leverage the strengths of both the global and local models. For example, Pan et al. (2010) proposed a network-based radiometric equalization method to achieve a global and local performance. This method first performs linear global processing using a path propagation based method, followed by nonlinear optimization through a combination of the multilinear relationships established by dividing each overlapping region into subregions. Yu et al. (2017) proposed an auto-adapting global-to-local color balancing approach, first performing global correction on multiple images based on block adjustment, and then establishing a global color distribution surface by adaptive image blocking to realize local RN for image blocks. The local RN step can also be considered as a kind of post-processing. Li et al. (2020a) proposed a method based on “jointly optimizing global and local color consistency”, dividing the image into multiple regions based on superpixel segmentation, imposing local constraints on each subregion, and global constraints on the subregions belonging to the same image, to perform global and local RN in a unified model. This method provides natural and smooth local radiometric transitions, but the overall tonal consistency is insufficient, and the method is very time-consuming. Zhang et al. (2022) proposed a block adjustment based global-to-local RN method, where the global RN eliminates the global radiometric differences, and two local strategies are further employed to remove the local residual radiometric differences using NDVI data. Hong et al. (2023) presented an approach in which a global optimization based model is applied to the chromaticity channel, and the block-based Wallis transform method is used to optimize the luminance and contrast, effectively reducing the textural differences. The combined models often achieve better global and local radiometric consistency than the global and local models, and represent an important future direction for RN. However, the existing approaches still struggle to achieve satisfactory results in terms of overall radiometric tonal consistency, as well as natural and smooth local radiometric transitions. Meanwhile, the spatial texture details of the image should be effectively preserved, and the computation time should be fast enough to satisfy the actual requirements.

In this paper, to address these challenges, we propose a two-stage approach that joints block adjustment and variational optimization for global and local RN. Firstly, a block adjustment based global RN (BAGRN) model is used to eliminate the global radiometric differences. The images are then rectangularly blocked, and a variational optimization-based local RN (VOLRN) model is flexibly applied to address the remaining local differences. As a result, overall tonal consistency, smooth local transitions, good gradient preservation, and high processing efficiency can be achieved.

The rest of this paper is organized as follows. Section II introduces the proposed approach, Section III presents the experiments and a discussion, and Section IV provides our conclusions.

2. Methodology

2.1. Overview

The proposed approach includes two stages: block adjustment based global RN (BAGRN) and variational optimization based local RN (VOLRN). The processing flow for global and local RN is shown in Fig. 1. The purpose of the proposed approach is to efficiently perform global and local RN to eliminate the overall radiometric tone differences between images and the local radiometric feature discontinuities and edge artifacts, to achieve overall radiometric tone consistency and smooth local radiometric transitions between multiple images, and to effectively preserve the textural details and spatial information. The input to the approach is a set of geometrically aligned images with dramatic radiometric differences $\mathscr{L}\{\boldsymbol{I}_i\}_{i=1}^{N}$, and the output is a set of radiometrically normalized images $\mathscr{T}\{\widehat{\boldsymbol{I}}_i\}_{i=1}^{N}$. Note that I_i is the source image, $\widehat{I}_i$ denotes the radiometrically normalized image, and N is the number of input/output images.

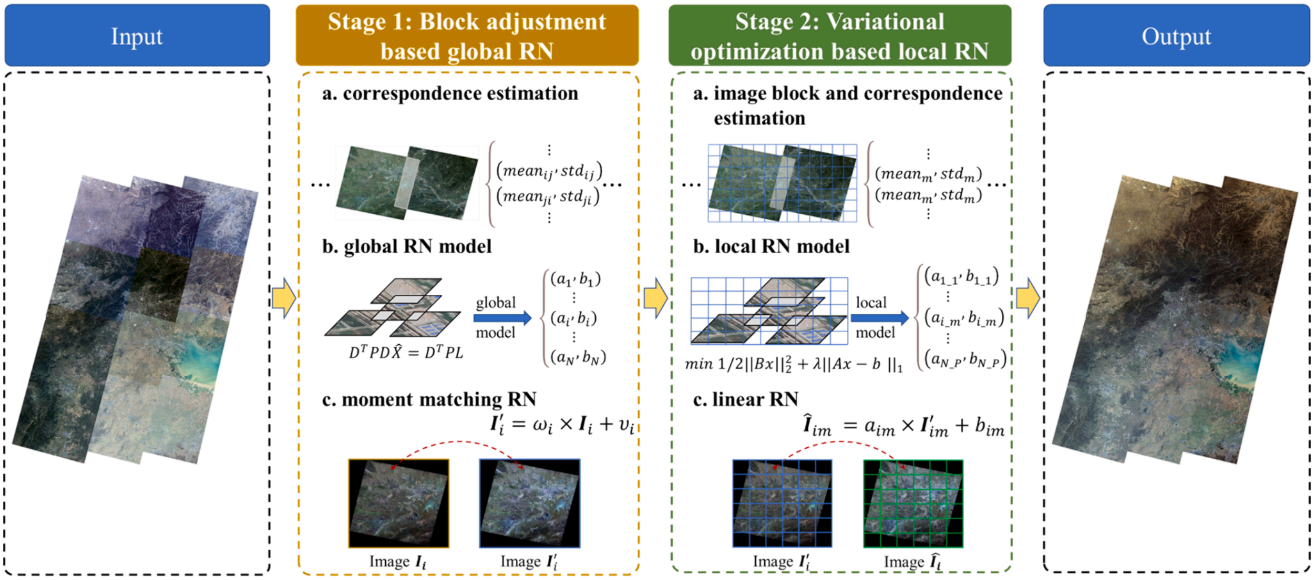


Fig. 1. Flowchart of global and local RN. Note that RN stands for radiometric normalization.

2.2. Block adjustment based global radiometric normalization (BAGRN)

The BAGRN model is designed to eliminate the overall radiometric tonal differences between images and reduce the cumulative errors caused by the transfer paths in the classical one-after-another pipeline. In this study, we built the BAGRN model based on the general idea of block adjustment RN (Zhang et al., 2022). The main processing flow is shown in Stage 1 of Fig. 1.

In the proposed approach, the mean μ and standard deviation δ of the image pixels are used as the radiometric correspondence. Given a set of geometrically aligned images $\mathscr{L}\{I_i\}_{i=1}^N$, there are a total of N frames, of which Q pairs of images have an overlapping relationship. All the pixels in the overlapping region of the image pairs can be extracted by the geographic location relationship, and the radiometric correspondences required for BAGRN model construction can be easily computed.

One of the basic assumptions of RN is that the radiometric information in the overlapping areas of image pairs should be equal or have minimal differences. The mean μ and standard deviation δ of an image are collectively referred to as the radiometric information of the image, and the purpose of RN is to obtain the compensation value $\theta(\theta_\mu, \theta_\delta)$ for the radiometric information of each image so that the radiometrically normalized image satisfies this assumption. Since the form of solving the compensation value of the mean μ and standard deviation δ is consistent, the following is an example of modeling the mean μ. The source i th and j th images with overlapping relationships can be represented by the radiometric information of their overlapping area as:

$$\theta_i - \theta_j = \mu_j - \mu_i \tag{1}$$

where μ_i, μ_j and θ_i , θ_j are the mean of the pixels and the compensated values in the overlapping area of the i th and j th images, respectively.

Q pairs of images with overlapping relationships can be listed as Q equations, as in Eq. (1), which can be expressed as:

$$\begin{pmatrix} \cdots & & & & & & & \\ & \cdots & & & & & & \\ & & \cdots & & & & & \\ 0 & \cdots & 1 & \cdots & 1 & \cdots & 0 & \\ & & & & \cdots & & & \\ & & & & & \cdots & & \\ & & & & & & \cdots & \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_j \\ \vdots \\ \theta_N \end{pmatrix} = \begin{pmatrix} \vdots \\ \vdots \\ \vdots \\ \mu_j - \mu_i \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \tag{2}$$

Eq. (2) can be abbreviated as:

$$D_\alpha X_\mu = L_\alpha \tag{3}$$

In Eq. (3), D_α is the overlapping matrix, where the elements of each row record a pair of overlapping images, and the column number of the two elements with a value of 1 is the identity number of the pair of images. D_α is the sparse matrix of $Q \times N$, X_μ is the vector of $N \times 1$, and L_α is the vector of $Q \times 1$.

For Eq. (3), it can be challenging when the number of images exceeds the number of overlapping regions, resulting in an underdetermined linear system. While the least squares method can obtain a solution that minimizes the sum of the squared residuals, such solutions can be sensitive to small variations in the input data or problem parameters. This sensitivity can cause the solution to vary significantly under different conditions. By introducing additional control conditions, the behavior of the solution can be stabilized and the occurrence of solution drift can be reduced.

Therefore, the solution can be obtained by specifying a certain image with a better radiometric tone as the control image, which does not need to be adjusted. We let the r th image be the control image, i.e., the mean compensation of this image is 0, which can be expressed as:

$$\theta_r = 0 \tag{4}$$

Eq. (4) can then be rewritten as:

$$\begin{pmatrix} 0 & \cdots & \cdots & 1 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \vdots \\ \theta_r \\ \vdots \\ \vdots \\ \theta_N \end{pmatrix} = (0) \tag{5}$$

Eq. (5) can be abbreviated as:

$$D_\beta X_\mu = L_\beta \tag{6}$$

In Eq. (6), D_β is the sparse matrix of $1 \times N$, and L_β is the zero vector of 1×1.

Combining Eq. (3) and Eq. (6), the BAGRN model can be constructed as:

$$DX_\mu = L \tag{7}$$

In Eq. (7), D is a sparse augmented matrix with dimension $(Q+1) \times N$, $D = \left(D_\alpha^T \; D_\beta^T\right)^T$, $L = \left(L_\alpha^T \; L_\beta^T\right)^T$, and L is the vector of $(Q+1) \times 1$.

In addition, the proportion of the pixel number in the overlapping area of the image is used as the weight to construct the weight matrix P of the least squares equation (Eq. (7)). P can be expressed as:

$$P = \begin{pmatrix} W & \boldsymbol{O} \\ \boldsymbol{O} & E \end{pmatrix} \tag{8}$$

In Eq. (8), $\boldsymbol{O}$ is the zero matrix, E is the unit matrix of 1×1 because there is only one image selected as the control image, and P and W are the diagonal matrices of $(Q+1) \times (Q+1)$ and $Q \times Q$, respectively. Each element of W represents the proportion of the pixel number in the i th pair of overlapping areas to the sum of the pixel number in all the overlapping areas.

The BAGRN model can be solved using the weighted least squares method, so that the compensation coefficients of the mean μ for all the images can be obtained:

$$\widehat{X}_\mu = \left(\theta_\mu^1 \; \theta_\mu^2 \; \cdots \; \theta_\mu^N\right)^T \tag{9}$$

Similarly, the compensation coefficients of the standard deviation δ can also be obtained.

After obtaining the compensation coefficients for the mean μ and standard deviation δ of each image, moment matching is applied to the RN to obtain a better effect. RN based on moment matching is achieved by approximating the moments of the data distribution of the target image to those of the reference image. By calculating the mean and standard deviation of the two above, corresponding approximate Gaussian distributions can be constructed, allowing moment matching to be performed to achieve radiometric correction. RN based on moment matching is calculated as:

$$f_k^{'} = \omega \times f_k + \upsilon \tag{10}$$

The image to be radiatively normalized is considered as the target image, and the image with an overlapping relationship with the target image is considered as the reference image. In Eq. (10), f_k represents the value of the k th pixel of the target image, and $f_k^{'}$ represents the value of the k th pixel of the target image after RN. The gain coefficient ω and offset coefficient υ for moment matching are calculated as follows:

$$\begin{cases} \omega = \dfrac{\delta_{ref}}{\delta_{tar}} \\ \upsilon = \mu_{ref} - \mu_{tar} \times \omega \end{cases} \tag{11}$$

where μ_{ref} and δ_{ref} respectively denote the mean and standard deviation of the pixels in the overlapping region of the reference image, while μ_{tar} and δ_{tar} respectively denote the mean and standard deviation of the pixels in the overlapping region of the target image.

2.3. Variational optimization based local radiometric normalization (VOLRN)

The VOLRN model is designed to eliminate the residual local radiometric feature discontinuities and edge artifacts after global RN, the model includes four steps: image blocking, variational model construction, variational model optimization, and weighted linear interpolation of the coefficients. The main processing flow is shown in Stage 2 of Fig. 1.

2.3.1. Image blocking

The first step is to generate the minimum outer rectangle of the image set $\mathscr{L}\{\boldsymbol{I}_i\}_{i=1}^N$ with the geographic grid information. From the geographic range of all the source images, the minimum outer rectangle of the image set can be calculated. Based on the geographic coordinates and the grid size, the location and geometric information of all the grid cells can be generated. The next step is to obtain valid image blocks and pairs of image blocks with overlapping relationships. Each grid cell is traversed to determine the image blocks it contains, and the numbering is incremented, starting from one. The image blocks are numbered if there are valid pixels; otherwise, they are not numbered. For valid image blocks that fall in the same grid cell, valid overlapping image block pairs are formed. After traversing all the grid cells, information about all the valid image blocks and pairs of image blocks with overlapping relationships can be generated, including the location of each image block and its position and size in the source image. Finally, the mean and standard deviation of the pixels of each image block can be calculated and the desired radiometric correspondences can be obtained. The specific procedure is shown in Fig. 2.

2.3.2. Variational model construction

For image RN, a widely accepted underlying assumption is that there is a linear relationship between the pixels in the overlapping areas of the images before and after RN, which can be expressed as:

$$f_k^{'} = a \times f_k + b \tag{12}$$

In Eq. (12), $f_k^{'}$ is the RN result of the k th pixel, f_k is the source pixel value of the k th pixel, a is the multiplicative coefficient, and b is the additive coefficient.

Assuming that the RN coefficients for any image block are a and b, and that the mean and the standard deviation of the radiometrically normalized image block are $\mu^{'}$ and $\delta^{'}$, this can be expressed as:

$$\begin{cases} \mu^{'} = a\mu + b \\ \delta^{'} = a\delta \end{cases} \tag{13}$$

From Eq. (13), the mean and standard deviation of the pixels of the radiometrically normalized image can be expressed as the mean and standard deviation of the pixels of the source image and the RN coefficients.

Based on the previous assumptions, the radiometric difference relationship between pairs of image blocks with overlapping relationships after RN can be constructed.

Suppose that the input image set $\mathscr{L}\{\boldsymbol{I}_k\}_{k=1}^N$ is divided into T image blocks according to the method described in Section 2.3.1, and there are M pairs of image blocks with an overlapping relationship. We let the i th image block and the j th image block have an overlapping relationship, and denote them as the t th $(t = 1, 2, \cdots, M)$ pair of overlapping image blocks. The mean and the standard deviation of the source pixels of the i

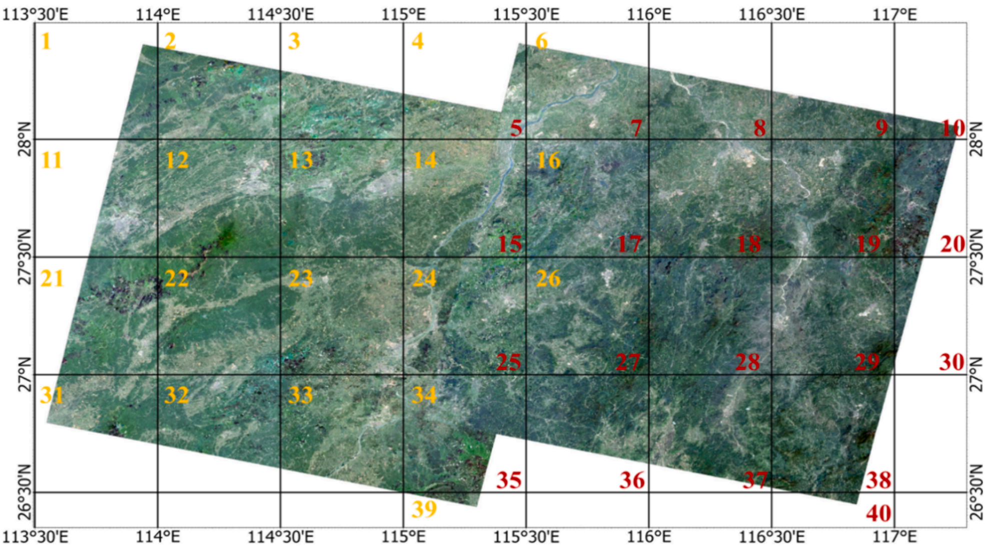


Fig. 2. Schematic diagram of image blocking and numbering.

th and j th image blocks are μ_i, μ_j and δ_i, δ_j, respectively. The RN coefficients of the i th and j th image blocks are a_i, b_i and a_j, b_j, respectively. Their mean and the standard deviation of the pixels of the RN results are μ_i', μ_j' and of δ_i', δ_j', respectively. Then, from Eq. (13), the mean and standard deviation of the pixels before and after RN have the following relationship:

$$\begin{cases} \mu_i' = a_i\mu_i + b_i \\ \delta_i' = a_i\delta_i - \delta_i \\ \mu_j' = a_j\mu_j + b_j \\ \delta_j' = a_j\delta_j - \delta_j \end{cases} \tag{14}$$

Therefore, the radiometric difference between the i th image block and the j th image block after RN can be expressed as:

$$\begin{cases} \varepsilon_{i,j}^{\mu} = \mu_i^{'} - \mu_j^{'} = a_i\mu_i + b_i - \left(a_j\mu_j + b_j\right) \\ \varepsilon_{i,j}^{\delta} = \delta_i^{'} - \delta_j^{'} = a_i\delta_i - a_j\delta_j \end{cases} \tag{15}$$

M pairs of image blocks with overlapping relationships can be listed as M equations, as in Eq. (15), which can be expressed as:

$$\begin{pmatrix} \vdots \\ \vdots \\ \varepsilon_{i,j}^{\mu} \\ \varepsilon_{i,j}^{\delta} \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & \cdots & \mu_i & 1 & \cdots & -\mu_j & -1 & \cdots & 0 \\ 0 & \cdots & \delta_i & \cdots & \cdots & -\delta_j & \cdots & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \vdots \\ \vdots \\ a_i \\ b_i \\ \vdots \\ \vdots \\ a_j \\ b_j \\ \vdots \\ \vdots \end{pmatrix} \tag{16}$$

Eq. (16) can then be abbreviated as:

$$\varepsilon = Bx \tag{17}$$

In Eq. (17):

$$\varepsilon = \left(\cdots \ \cdots \ \varepsilon_{i,j}^{\mu} \ \varepsilon_{i,j}^{\delta} \ \cdots \ \cdots\right)^{\mathrm{T}} \tag{18}$$

$$x = \left(\cdots \ \cdots \ a_i \ b_i \ \cdots \ a_j \ b_j \ \cdots \ \cdots\right)^{T} \tag{19}$$

$$B = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & \cdots & \mu_i & 1 & \cdots & -\mu_j & -1 & \cdots & 0 \\ 0 & \cdots & \delta_i & \cdots & \cdots & -\delta_j & \cdots & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \tag{20}$$

In Eq. (17), B is the coefficient matrix of $2M \times 2T$, x is a column vector of $2T \times 1$, and ε is a column vector of $2M \times 1$.

In addition, the textural details and spatial information of the image before and after RN should be kept as consistent as possible, so that the radiometric difference relationship of the image before and after RN can be constructed.

We let the mean and standard deviation of the source pixels of the k th $(k = 1, 2, \cdots, T)$ image block be μ_k, δ_k, the RN coefficients of the image block be a_k, b_k, and the mean and standard deviation of the pixels of the image block after RN be μ_k', δ_k'. Then, from Eq. (13), the mean and standard deviation of the pixels of the image block before and after the RN have the following relationship:

$$\begin{cases} \mu_k' = a_k\mu_k + b_k \\ \delta_k' = a_k\delta_k \end{cases} \tag{21}$$

Therefore, the radiometric difference between the k th image block before and after RN can be expressed as:

$$\begin{cases} \tau_k^{\mu} = \mu_k^{'} - \mu_k = a_k\mu_k + b_k - \mu_k \\ \tau_k^{\delta} = \delta_k^{'} - \delta_k = a_k\delta_k - \delta_k \end{cases} \tag{22}$$

T image blocks can be listed as T equations, as in Eq. (22), which can be expressed as:

$$\begin{pmatrix} \vdots \\ \vdots \\ \tau_k^{\mu} \\ \tau_k^{\delta} \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \mu_k & 1 & \cdots & 0 \\ 0 & \cdots & \delta_k & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \vdots \\ \vdots \\ a_k \\ b_k \\ \vdots \\ \vdots \end{pmatrix} - \begin{pmatrix} \vdots \\ \vdots \\ \mu_k \\ \delta_k \\ \vdots \\ \vdots \end{pmatrix} \tag{23}$$

Eq. (23) can then be abbreviated as:

$$\tau = Ax - b \tag{24}$$

In Eq. (23): $\tau = (\cdots\ \cdots\ \tau_k^{\mu}\ \tau_k^{\delta}\ \cdots\ \cdots)^{\mathrm{T}}$ (25)

$$x = (\cdots\ \cdots\ a_k\ b_k\ \cdots\ \cdots)^T \tag{26}$$

$$b = (\cdots\ \cdots\ \mu_k\ \delta_k\ \cdots\ \cdots)^T \tag{27}$$

$$A = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \mu_k & 1 & \cdots & 0 \\ 0 & \cdots & \delta_k & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \tag{28}$$

In Eq. (24), A is the coefficient matrix of $2T \times 2T$, and τ, x, b are the column vectors of $2T \times 1$.

Based on the previously constructed radiometric difference relationship of image block pairs with overlapping relationships after RN and the radiometric difference relationship of image blocks before and after RN, the VOLRN model can be constructed. For the radiometric difference of image block pairs with overlapping relationships after RN, the widely used ℓ_2 norm constraint is imposed on them so that the sum of squares of their radiometric differences is minimal. As for the radiometric difference of image blocks before and after RN, the previous related studies also generally imposed an ℓ_2 norm constraint on them. For example, Yu et al. (2017) and Li et al. (2020a) applied an ℓ_2 norm constraint to this radiometric difference term and solved it using the least squares method to minimize the sum of squares of the radiometric difference. However, when the image radiometric difference is dramatic, applying the ℓ_2 norm constraint to this term is often not an optimal way to form a reliable constraint and may even weaken the effect of RN. After application of the BAGRN model, there will already be good overall radiometric tonal consistency between images, and in the VOLRN stage, the main objective is to eliminate the local radiometric differences and edge artifacts remaining after the global RN. For the local radiometric differences after BAGRN model application, we consider that the image blocks that are not in the overlapping area remain unchanged or change very little, while the image blocks in the overlapping area and the transition between the overlapping and non-overlapping areas should be radiometrically normalized effectively. Therefore, we propose a new idea to impose an ℓ_1 norm constraint on this radiometric difference term, because the ℓ_1 norm has better sparsity, which can be used to radiometrically normalize only some of the image blocks, and not the other blocks, i.e., to keep the RN coefficients of some of them to be $a_k = 1, b_k = 0$. Thus, the final VOLRN model can be constructed as:

$$E(x) = \frac{1}{2}\|Bx\|_2^2 + \lambda\|Ax - b\|_1 \tag{29}$$

In Eq. (29), the energy functional $E(x)$ is the objective function, λ is the regularization parameter, and $\|\bullet\|_1$ and $\|\bullet\|_2$ denote the ℓ_1 norm and ℓ_2 norm, respectively.

2.3.3. Variational model optimization

Clearly, Eq. (29) is an unconstrained optimization problem with both ℓ_2 norm and ℓ_1 norm, which is difficult to convert into a linear problem for solving, but can be solved by using the alternating direction method of multipliers (ADMM) (Boyd, 2010). The problem can be rewritten in ADMM form as:

$$\begin{aligned} & min\frac{1}{2}\|Bx\|_2^2 + \lambda\|z\|_1 \\ & s.t. Ax - z - b = 0 \end{aligned} \tag{30}$$

For which the augmented Lagrangian function is:

$$L_\rho(x,z,y) = \frac{1}{2}\|Bx\|_2^2 + \lambda\|z\|_1 + \frac{\rho}{2}\|Ax - z - b + u\|_2^2 \tag{31}$$

An optimization formulation for each subproblem can be obtained as follows:

$$\begin{cases} x^{k+1} := argmin_x L_\rho\left(x, z^k, u^k\right) = argmin_x\left(\frac{1}{2}\|Bx\|_2^2 + \frac{\rho}{2}\|Ax - z - b + u\|_2^2\right) \\ z^{k+1} := argmin_z L_\rho\left(x^{k+1}, z, u^k\right) = argmin_z\left(\lambda\|z\|_1 + \frac{\rho}{2}\|Ax - z - b + u\|_2^2\right) \\ u^{k+1} := u^k + Ax^{k+1} - z^{k+1} - b \end{cases} \tag{32}$$

The solution formula for each subproblem can be obtained by the derivative method, which is calculated analytically as:

$$\begin{cases} x^{k+1} := \left(B^T B + \rho A^T A\right)^{-1}\left(\rho A^T\left(z^k - u^k + b\right)\right) \\ z^{k+1} := S_{\lambda/\rho}\left(Bx^{k+1} + u^k - b\right) \\ u^{k+1} := u^k + Ax^{k+1} - z^{k+1} - b \end{cases} \tag{33}$$

To solve each subproblem in Eq. (33), the z subproblem can be solved using the soft threshold method, while for the x subproblem, due to the large scale of the equations, the preconditioned conjugate gradient method is used for the solution. In the specific solution process, each subproblem can be solved by iteratively alternating until a preset number of iterations or iteration errors is reached, where the RN coefficients for each image block can be obtained.

2.3.4. Weighted linear interpolation of coefficients

After application of the VOLRN model, T pairs of RN coefficients corresponding to T image blocks can be obtained. If the coefficients of each image block are directly used to radiometrically normalize all the pixels within the block, the result may suffer from the block effect, leading to a significant reduction in data availability. Therefore, the nine pairs of RN coefficients of the image block and its eight neighboring image blocks are used for inverse distance weighted linear interpolation to calculate the RN coefficients for each pixel. The interpolation principle is shown in Fig. 3, where the red dot represents the center of the image block (m, n), the blue dots represent the center of the eight neighboring image blocks of image block (m,n), and the small black dot represents a pixel f_k falling in image block (m, n). The inverse of the distance is used as the weight of the linear interpolation, and the farther the pixel to be solved is from the center position of a block, the less the block affects it, and vice versa. The RN coefficient for each pixel can be calculated as shown in Eqs. (34)–(35):

$$\begin{cases} a_k = \overrightarrow{a}_{m,n} \times \overrightarrow{C}_{m,n} \\ b_k = \overrightarrow{b}_{m,n} \times \overrightarrow{C}_{m,n} \end{cases} \tag{34}$$

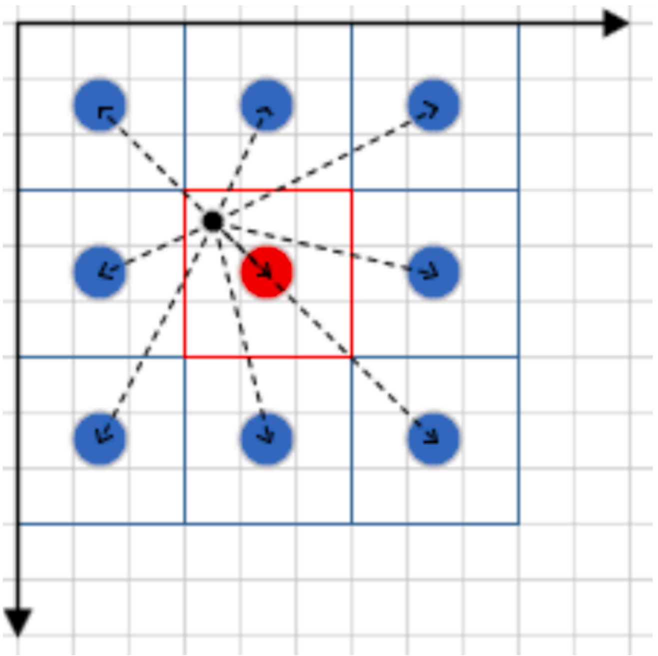


Fig. 3. Schematic of inverse distance weighted linear interpolation of parameters.

In Eq. (34), a_k, b_k represent the final RN coefficients of pixel f_k; $\overrightarrow{a}_{m,n}$, $\overrightarrow{b}_{m,n}$ represent the coefficients of image block (m,n) where pixel f_k is located and its eight neighboring image blocks; and $\overrightarrow{C}_{m,n}$ represents the inverse of the distances from pixel f_k to the centers of the nine image blocks, which serve as the weights of the coefficients, and can be expressed as:

$$\left\{\begin{array}{r}\overrightarrow{a}_{m,n} = \left(a_{m,n}, a_{m\pm1,n}, a_{m,n\pm1}, a_{m\pm1,n\pm1}\right)\\ \overrightarrow{b}_{m,n} = \left(b_{m,n}, b_{m\pm1,n}, b_{m,n\pm1}, b_{m\pm1,n\pm1}\right)\\ \overrightarrow{C}_{m,n} = \left(1/C_{m,n}, 1/C_{m\pm1,n}, 1/C_{m,n\pm1}, 1/C_{m\pm1,n\pm1}\right)\end{array}\right. \tag{35}$$

After obtaining the RN coefficients a_k, b_k for each pixel f_k, a linear calculation can be applied to each pixel using Eq.(12).

3. Experimental results and discussion

We conducted a series of experiments to validate the effectiveness, robustness, and efficiency of the proposed approach on five remote sensing datasets from different platforms, all with different spatial resolutions, and all suffering from dramatic radiometric differences. The key information of the five datasets is listed in Table 1. Note that dataset #4 is a subset of dataset #5, which was selected for the parameter analysis. Comparative evaluation experiments with three state-of-the-art approaches were first performed using large datasets #2, #3, and #5, followed by hyperparametric analysis experiments using the relatively small datasets #1, #3, and #4. Finally, ablation experiments using datasets #2, #3, and #5 were conducted to compare and evaluate the performance and difference in the results of the proposed two-stage model approach. Both the proposed approach and the comparison approaches were implemented in C++ on Windows 10, and all the experiments were performed on a computer with an Intel i9-10850 k CPU (3.60 GHz) and 64 GB of RAM.

3.1. Evaluation metrics

To quantitatively evaluate the effectiveness of the RN approaches, four evaluation metrics are adopted here. The radiometric differences between adjacent images are characterized by utilizing the metrics of the absolute difference of the mean (*ADM*), the absolute difference of the standard deviation (*ADSD*) (Zhang et al., 2022), and the color distance (*CD*) of the overlapped areas of images. The gradient loss (*GL*) (Xia et al., 2019) represents the gradient loss before and after RN. Meanwhile, to comprehensively evaluate the radiometric difference, we define the radiometric difference in the overlapping area (*RDOA*) as the average of the *ADM*, *ADSD*, and *CD*. And *Ave* is defined as the average of the *ADM*, *ADSD*, *CD*, and *GL*. Lower values for these metrics represent better RN and gradient preservation capacity. The metrics are calculated as follows:

$$ADM = \sum\nolimits_{I_i \cap I_j \neq \varnothing} \frac{\left|\Delta M\left(\widehat{I}_{ij}, \widehat{I}_{ji}\right)\right|}{N_l} \tag{36}$$

$$ADSD = \sum\nolimits_{I_i \cap I_j \neq \varnothing} \frac{\left|\Delta S\left(\widehat{I}_{ij}, \widehat{I}_{ji}\right)\right|}{N_l} \tag{37}$$

$$CD = \sum\nolimits_{I_i \cap I_j \neq \varnothing} w_{ij} \frac{\Delta H\left(\widehat{I}_{ij}, \widehat{I}_{ji}\right)}{N_b} \tag{38}$$

$$GL = \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\Delta G\left(I_i, \widehat{I}_i\right)}{N_p} \tag{39}$$

$$RDOA = (ADM + ADSD + CD)/3 \tag{40}$$

$$Ave = (ADM + ADSD + CD + GL)/4 \tag{41}$$

where I_i is the source image, $\widehat{I}_i$ denotes the radiometrically normalized image, and $\widehat{I}_{ij}$ represents the area of $\widehat{I}_i$ overlapped by $\widehat{I}_j$. The normalized weight w_{ij} is set proportional to the area of $\widehat{I}_{ij}\left(\sum \overline{w}_{ij} = 1\right)$. Moreover, $\Delta M(\bullet)$ computes the difference of the mean, $\Delta S(\bullet)$ computes the difference of the standard deviation, $\Delta H(\bullet)$ computes the difference between histograms by bins, and $\Delta G(\bullet)$ computes the difference between the gradient orientation maps by pixels. N is the number of images, N_l is the number of overlapping areas of the images, N_b is the bin number of a histogram, and N_p is the pixel number of I_i.

3.2. Comparative evaluation

The proposed approach was compared with the three state-of-the-art approaches of the method developed by (Xia et al., 2019), the method developed by (Zhang et al., 2022), and Inpho OrthoVista.[1] Inpho OrthoVista is well-known and widely used commercial software, which shows a good performance in RN and mosaicking of remote sensing images. In the following, we refer to these three approaches as CDFGA (Cumulative Distribution Function based Global Adjustment), SDGLBA (Statistical Distribution based Global and Local Block Adjustment) and GTA (Global Tilting Adjustment). Since all of the above approaches, except GTA, have difficulty in handling large datasets, we rewrote and optimized them using C/C++ to improve their data processing capabilities. For a fair comparison, we made sure that the RN results before and after optimization of the program did not change, and the relevant parameters were set according to the author's recommendations. Detailed descriptions of all the comparison approaches are provided in Table 2. Note that, in the comparison experiments, the regularization parameter was set to 0.5, and the image block sizes were set to 200, 400, and 600, corresponding to datasets #2, #3, and #5 with spatial

---

[1] https://geospatial.trimble.com/en/products/software/trimble-inpho.

Table 1
The details of the five datasets used in the experiments.

Dataset	Image number	Image mean size	Spatial resolution	Mean overlap	Main feature types	Data volume (GB)	Covered area (km^2)	Platform
#1	12	7765 × 7890	30 m	14 %	city, mountain, vegetation	2.05	307,987	Landsat 8
#2	24	8017 × 7121	30 m	12 %	city, water, mountain, vegetation	5.65	784,380	Landsat 8
#3	8	21546 × 16062	16 m	16 %	desert, mountain	7.76	283,754	GF-1/6, HJ2A, HJ2B
#4	8	32762 × 30213	1 m	23 %	city, water, mountain	22.8	4,884	GF-1/2/6/7
#5	39	30938 × 27179	1 m	13 %	city, water, mountain, vegetation	93.3	17,120	GF-1/2/6/7

Table 2
Detailed introduction to the three comparison approaches used in the experiments.

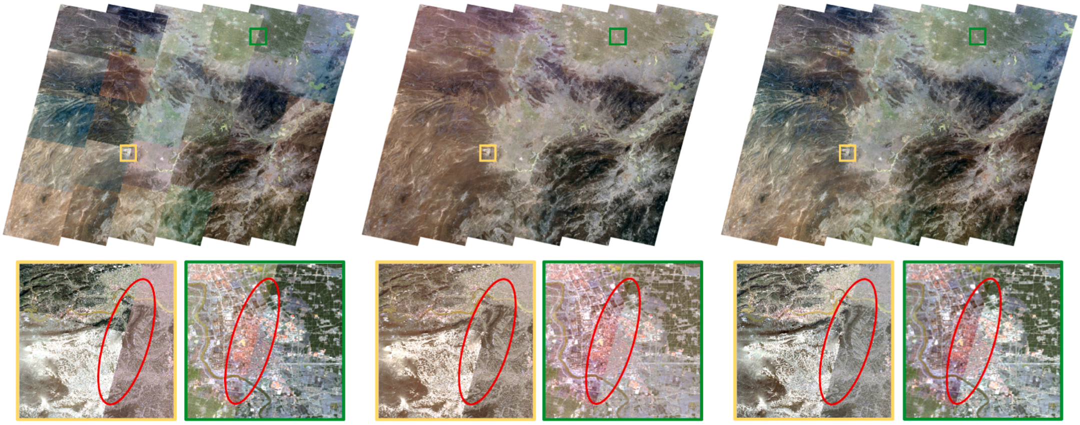
Approach	Abbreviation	Model	Correspondence	Optimization solver	Implementation
(Xia et al., 2019)	CDFGA	Global	Same quantiles in cumulative distribution functions	Constrained linear least squares	C/C++
Inpho OrthoVista	GTA	Global	—	—	—
(Zhang et al., 2022)	SDGLBA	Global-local	Mean and standard deviation	Weighted least squares	C/C++

resolutions of 30 m, 16 m, and 1 m, respectively. An analysis of the parameter settings is provided in Section 3.3.

3.2.1. Visual judgment

We conducted experiments and evaluations on datasets #2, #3, and #5, and the RN results for each dataset show the global and local RN effects of the different approaches. It is worth noting that the yellow and green boxes labeled in the full map correspond to the two local zoomed-in maps below it, and the red elliptical circles in the local zoomed-in map mark the main radiometric differences that exist.

We firstly conducted experiments on dataset #2, which is made up of 30 multi-temporal images acquired by the Landsat 8 satellite with a spatial resolution of 30 m. Fig. 4**(a)** shows the input images, where it can be seen that there are very large radiometric differences between the images. In terms of the global RN effect, the results of all the approaches have good overall radiometric tonal consistency. However, the results of CDFGA, GTA, BAGRN, and SDGLBA still have some radiometric differences at the image intersections on the left side, especially the results of GTA, which have obvious radiometric differences in the transition area between the mountains and the plains in the middle of the area. The RN


(a) Input images (b) CDFGA (c) GTA

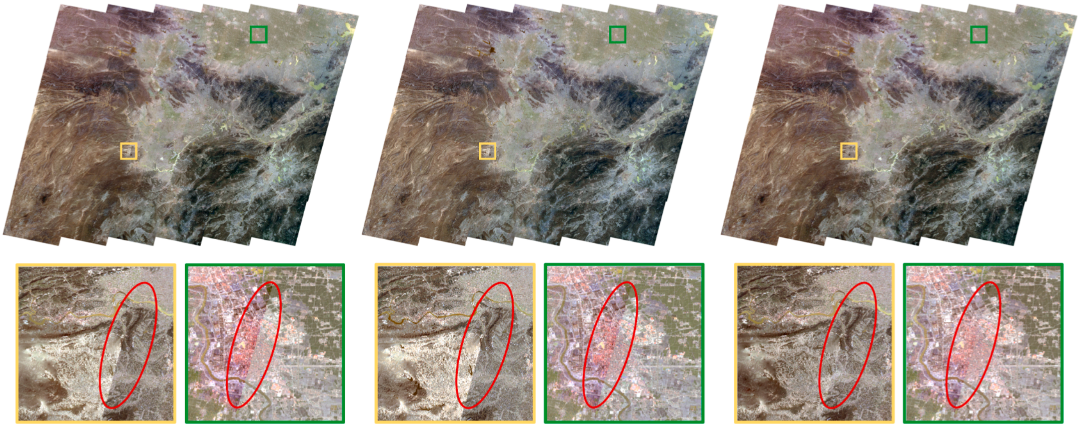

(d) BAGRN (e) SDGLBA (f) Proposed

Fig. 4. The RN results for dataset #2. (a) The input images. (b)–(f) The results of CDFGA, GTA, BAGRN, SDGLBA, and the proposed approach, respectively.

results of the comparison methods are shown in Fig. 4**(b)–(e)**, respectively. In contrast, the results of the proposed approach shown in Fig. 4 **(f)** effectively eliminate the radiometric differences at the image intersections. Regarding the local RN effect, the first zoomed-in map shows the RN results for the mountainous area, which is the intersection of the four images, where obvious local radiometric differences can be seen in the results of the other approaches. In contrast, the proposed approach performs very well, with no radiometric differences visible to the naked eye. The second zoomed-in map shows the RN results for the urban area, where the results of GTA are greenish and show obvious radiometric differences, and the results of CDFGA, BAGRN, and SDGLBA also show some radiometric differences, while the results of the proposed approach show very small radiometric differences. From Fig. 4, it can be seen that, from the experimental results obtained on dataset #2, the local RN effect of the proposed approach, and especially the smoothness of the transitions, is much better than that of the other approaches, while maintaining the overall radiometric tonal consistency.

We then conducted experiments on dataset #3, which consists of 8 multi-temporal images acquired by the Chinese Gaofen-1/6 and HJ2A/B satellites with a spatial resolution of 16 m. Fig. 5**(a)** shows the input images, where the intersections of the first image on the left, the four in the middle, and the three on the right all show severe radiometric differences. In terms of the global RN effect, the results of CDFGA in Fig. 5 **(b)** show vivid hues and good tonal consistency, but some areas are bright and distorted, and there are obvious radiometric differences at the intersections of the multiple images. The results of GTA shown in Fig. 5**(c)** show obvious tonal inconsistencies, with the four images on the left and the four images on the right showing two obvious color tones, and obvious radiometric differences can be seen at each image intersection, which is probably due to the higher contrast and saturation. The results of BAGRN, SDGLBA, and the proposed approach shown in Fig. 5 **(d)–(f)**, respectively, show very good overall radiometric tonal consistency with good radiometric balance and moderate brightness and contrast. However, the results of BAGRN and SDGLBA also show radiometric differences at some image intersections, while the results of the proposed approach show no visible radiometric differences at image intersections. Regarding the local RN effect, the first and second zoomed-in maps show the RN effect for a desert region and for a mountainous region, respectively. The results of both CDFGA and GTA appear more vivid, with higher brightness and contrast, but both suffer from local radiometric differences and distortions. The results of BAGRN and SDGLBA show a natural transition with only slight radiometric differences, while the results of the proposed approach are radiometrically balanced with a smooth and natural transition and no local radiometric differences. From Fig. 5, it can be seen that the experimental results of the proposed approach on dataset #3 show the best RN effect.

Finally, we conducted experiments on dataset #5, which consists of 39 multi-temporal images acquired by the Chinese Gaofen-1/2/6/7 satellites with a spatial resolution of 1 m. The dataset was preprocessed with geometric registration, pansharpening, super-resolution reconstruction, and cloud removal, according to our previous work (Lin et al., 2024). Fig. 6**(a)** shows the input image, where there are again very dramatic radiometric differences between the images. In terms of the global RN effect, the results of GTA shown in Fig. 6**(c)** show the worst performance, with obvious tonal inconsistencies in the left, upper right, and lower regions, and obvious radiometric differences at the image intersections between the left and lower right regions, with higher brightness and saturation. The results of CDFGA, BAGRN, SDGLBA, and the proposed approach shown in Fig. 6 (b), (d)–(f), respectively, show better radiometric tonal consistency, with no obvious radiometric differences at the image intersections, but the results of CDFGA still show brighter and more distorted effects. Regarding the local RN effect, the first zoomed-in map shows the RN effect in the area where the water body meets the island, which is the intersection of the four images. The results of CDFGA and GTA show obvious unsmooth radiometric transitions at the image intersections. The results of SDGLBA are better than those of the first three approaches, while there is still a slight radiometric difference. The results of the proposed approach show smooth and natural transitions and no local radiometric differences. The second zoomed-in map shows the RN effect for the area where the mountain meets the countryside. Similarly, the RN results of CDFGA, GTA, and BAGRN show obvious local radiometric differences at the image intersections, and the results of GTA are darker. The results of SDGLBA also show small radiometric differences, while the results of the proposed approach are very pleasing, with no visible local radiometric differences, and the transitions are very smooth and natural. From the experimental results obtained on dataset #5 shown in Fig. 6, it is apparent that the local RN capacity of the proposed approach is much better than that of the other approaches, especially for the water and mountainous regions.

In summary, the RN results of the proposed approach on the three datasets show a superior visual effect, and in terms of the global RN effect, it shows good radiometric tonal consistency, natural radiometric tone, and balanced brightness and contrast. Regarding the local RN effect, the radiometric transitions are smooth and natural, the details are clear and bright, and there are almost no visible local radiometric differences in most of the zoomed-in maps, which represents a great advantage over the other approaches. Therefore, it can be said that the proposed approach has very good global and local RN capabilities.

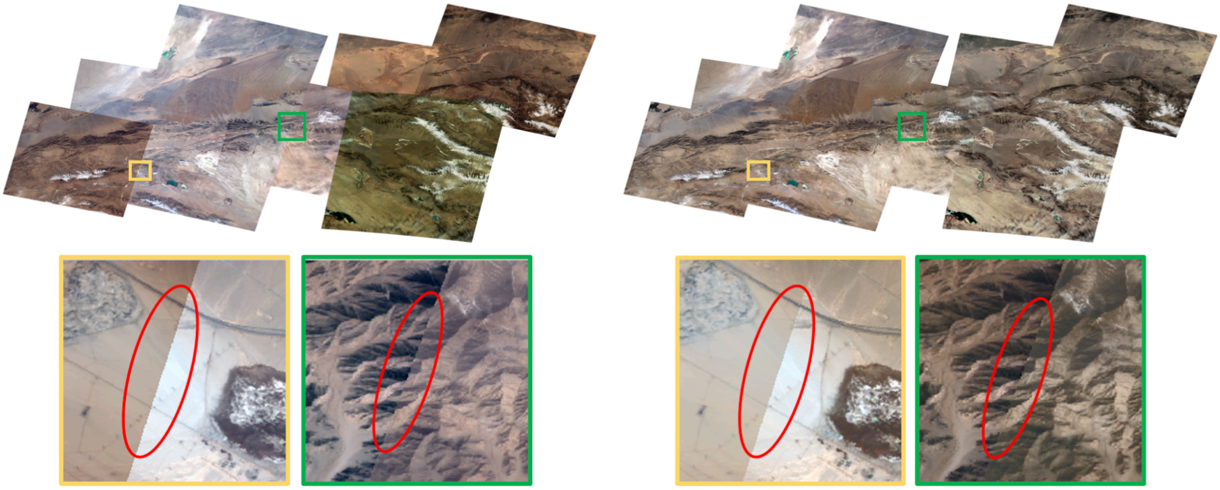
3.2.2. Quantitative evaluation

To convincingly demonstrate the superiority of the proposed approach, we performed a quantitative evaluation of all the approaches on datasets #2, #3, and #5 using the metrics introduced in Section 3.1. The results are presented in Table 3.

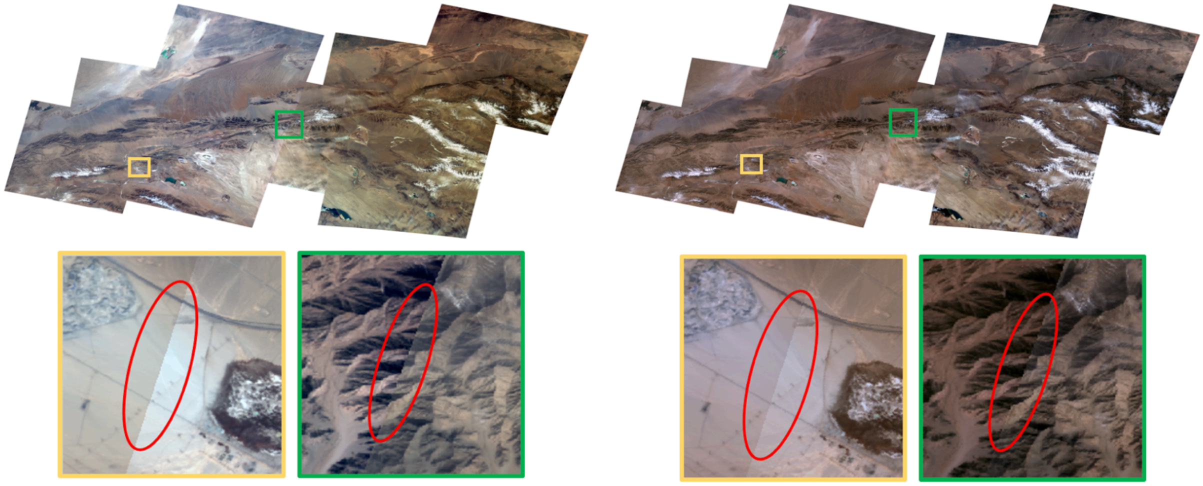
From Table 3, it can be seen that the *ADM*, *ADSD*, and *CD* values of the proposed approach are the lowest across all the datasets. Notably, these values are significantly lower than those of the other comparison approaches, indicating an excellent RN capacity and effective elimination of both the global and local radiometric differences. Specifically, optimal metric values are achieved on dataset #2, except for the BAGRN approach. In addition, the *GL* values achieved by the proposed approach, although not the lowest, are very close to the optimal values. Therefore, it can be considered that the proposed approach has a good gradient preservation capability. This is primarily because the RN models used in the two-stage approach are linear, thus avoiding significant loss of textural details of the images. On average, the metric values of the proposed approach across all the datasets are significantly lower than those of the other approaches, demonstrating a superior RN and gradient preservation performance.

In summary, the results of the qualitative and quantitative evaluations on the three datasets with different characteristics strongly demonstrate that the proposed approach has very good RN and gradient preservation capabilities, and is far superior to the other comparison approaches.

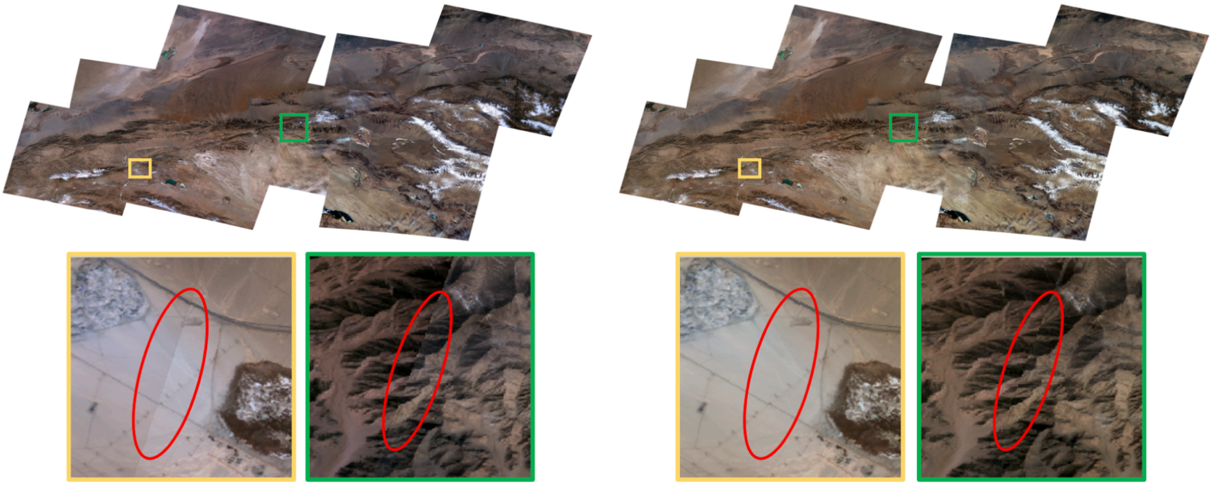
In addition, the computational times of all the approaches are reported in Table 3. The computational time is denoted by *Time*, and the units are m for minutes and h for hours. From Table 3, it can be seen that the proposed approach has the lowest computational time and the highest computational efficiency on all the datasets. The main reason for this is that the proposed approach uses a linear model, the image block method is simple, the model solving is fast, and the algorithm has good parallelism. Meanwhile, the computational time of the proposed approach when compared to BAGRN, i.e., with the addition of the VOLRN model, does not become much larger. This is because, with the two-stage model approach, some computational tasks can be merged in the specific programming implementations, especially the task of reading and writing the images, and thus the proposed



(a) Input images (b) CDFGA



(c) GTA (d) BAGRN



(e) SDGLBA (f) Proposed

Fig. 5. The RN results for dataset #3. (a) The input images. (b)–(f) The results of CDFGA, GTA, BAGRN, SDGLBA, and the.

(a) Input images (b) CDFGA (c) GTA

(d) BAGRN (e) SDGLBA (f) Proposed

Fig. 6. The RN results for dataset #5. (a) The input images. (b)–(f) The results of CDFGA, GTA, BAGRN, SDGLBA, and the proposed approach, respectively.

Table 3
Quantitative assessment of the RN results generated by the different approaches.

Dataset	Metric	Input	CDFGA	GTA	SDGLBA	BAGRN	Proposed
#2	*ADM*	11.692	2.893	1.172	2.797	3.363	**0.239**
	ADSD	3.312	1.997	2.955	1.611	1.727	**0.187**
	CD	7.461	1.989	1.871	2.012	2.339	**0.587**
	GL	0	0.182	1.222	0.311	**0.149**	0.159
	Ave	—	1.765	1.805	1.683	1.895	**0.293**
	Time	—	6.850 m	18.250 m	6.067 m	2.000 m	**3.733 m**
#3	*ADM*	19.130	3.795	1.677	2.173	2.009	**0.229**
	ADSD	3.959	2.734	3.374	2.026	2.120	**0.277**
	CD	11.577	2.609	2.477	2.255	2.313	**0.788**
	GL	0	0.358	0.926	0.344	**0.333**	0.375
	Ave	—	2.374	2.114	1.700	1.694	**0.417**
	Time	—	8.850 m	39.140 m	3.817 m	1.883 m	**3.267 m**
#5	*ADM*	19.349	2.609	3.890	1.256	1.710	**0.266**
	ADSD	3.257	1.874	3.204	0.541	0.623	**0.164**
	CD	10.770	1.948	3.410	1.046	1.374	**0.424**
	GL	0	**0.246**	0.901	0.336	0.282	0.289
	Ave	—	1.669	2.851	0.795	0.997	**0.286**
	Time	—	2.503 h	11.806 h	3.947 h	1.021 h	**1.506 h**

approach has a high processing efficiency. It can also be seen that CDFGA, SDGLBA, and BAGRN have the same magnitude of time as the proposed approach, although the computational time is longer, while the processing time of GTA is an order of magnitude higher, especially for the large dataset #5. In summary, the performance of the proposed approach is superior in terms of the computational efficiency.

3.3. Parameter analysis

In the VOLRN model, two key parameters need to be tuned: the regularization parameter λ and the image block size. In this section, we analyze the influence of different parameter settings on the RN results through quantitative assessment and visual evaluation.

3.3.1. Regulation parameter λ

Regularization parameter λ is designed to balance the radiometric consistency and data fidelity. In this experiment, dataset #1 was used for a comprehensive evaluation of the influence of this parameter. The results are shown in Fig. 7 and Fig. 8.

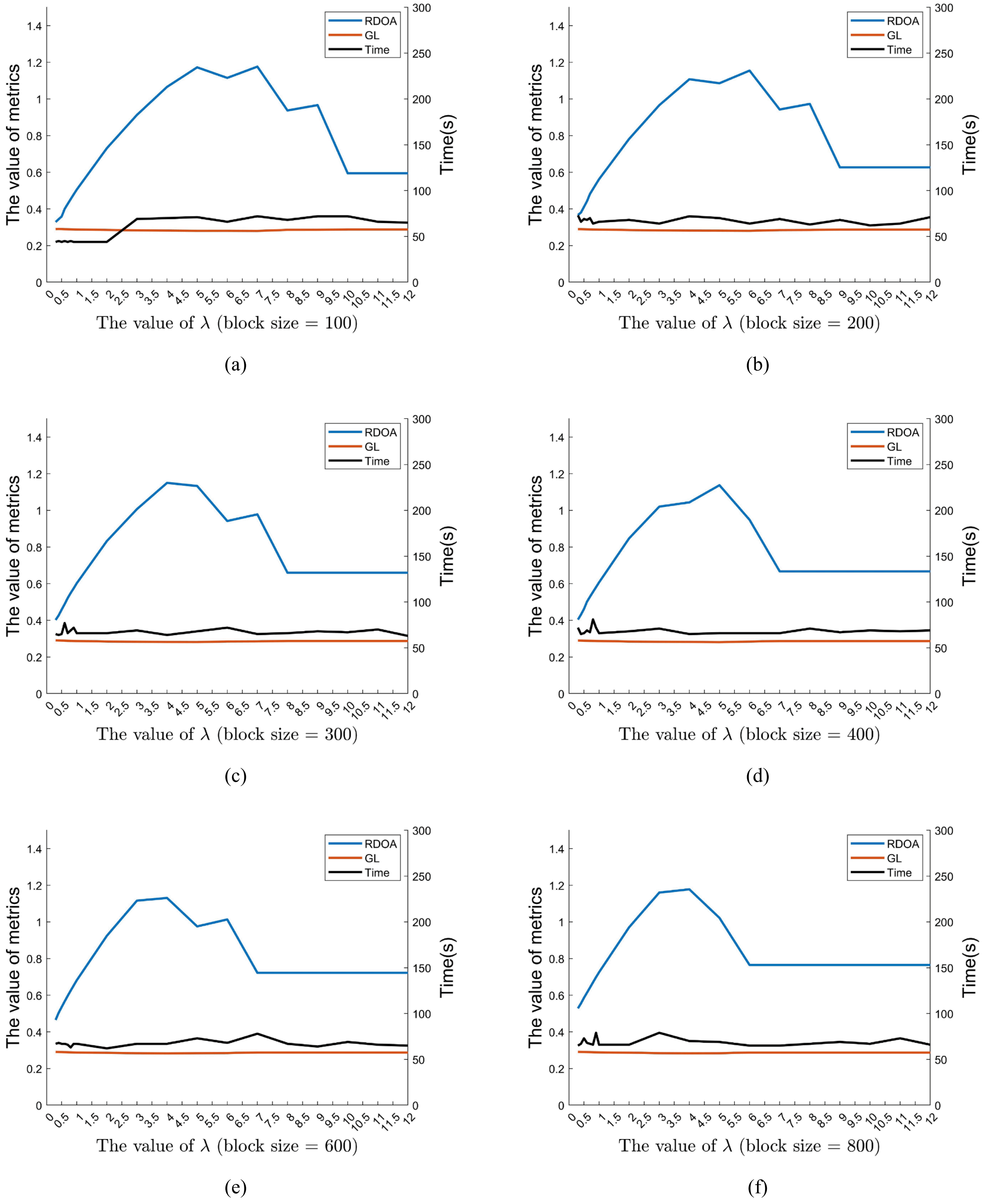


Fig. 7. Illustration of the influence of regulation parameter λ.

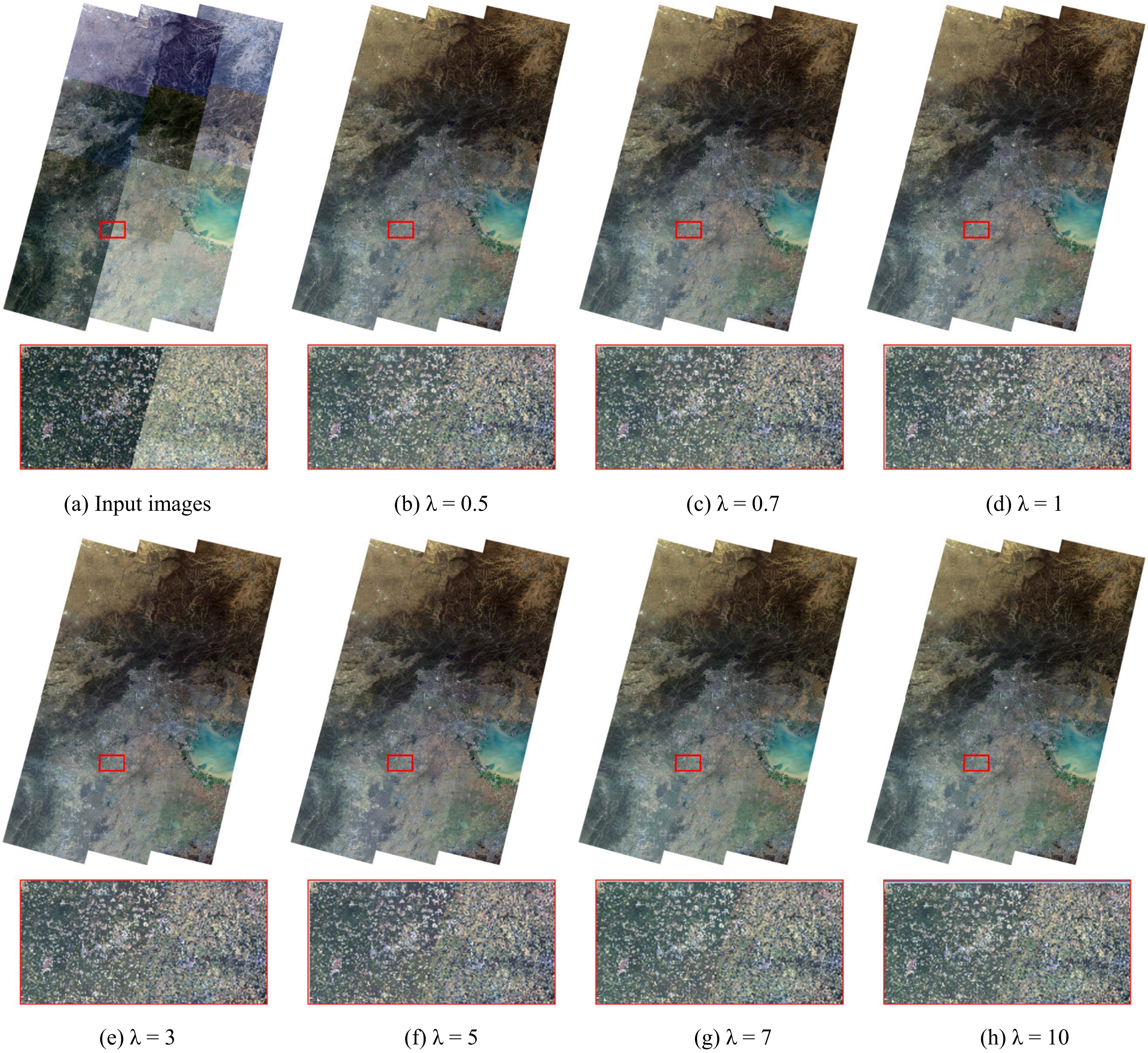


Fig. 8. Visual illustration of the influence of regulation parameter λ.

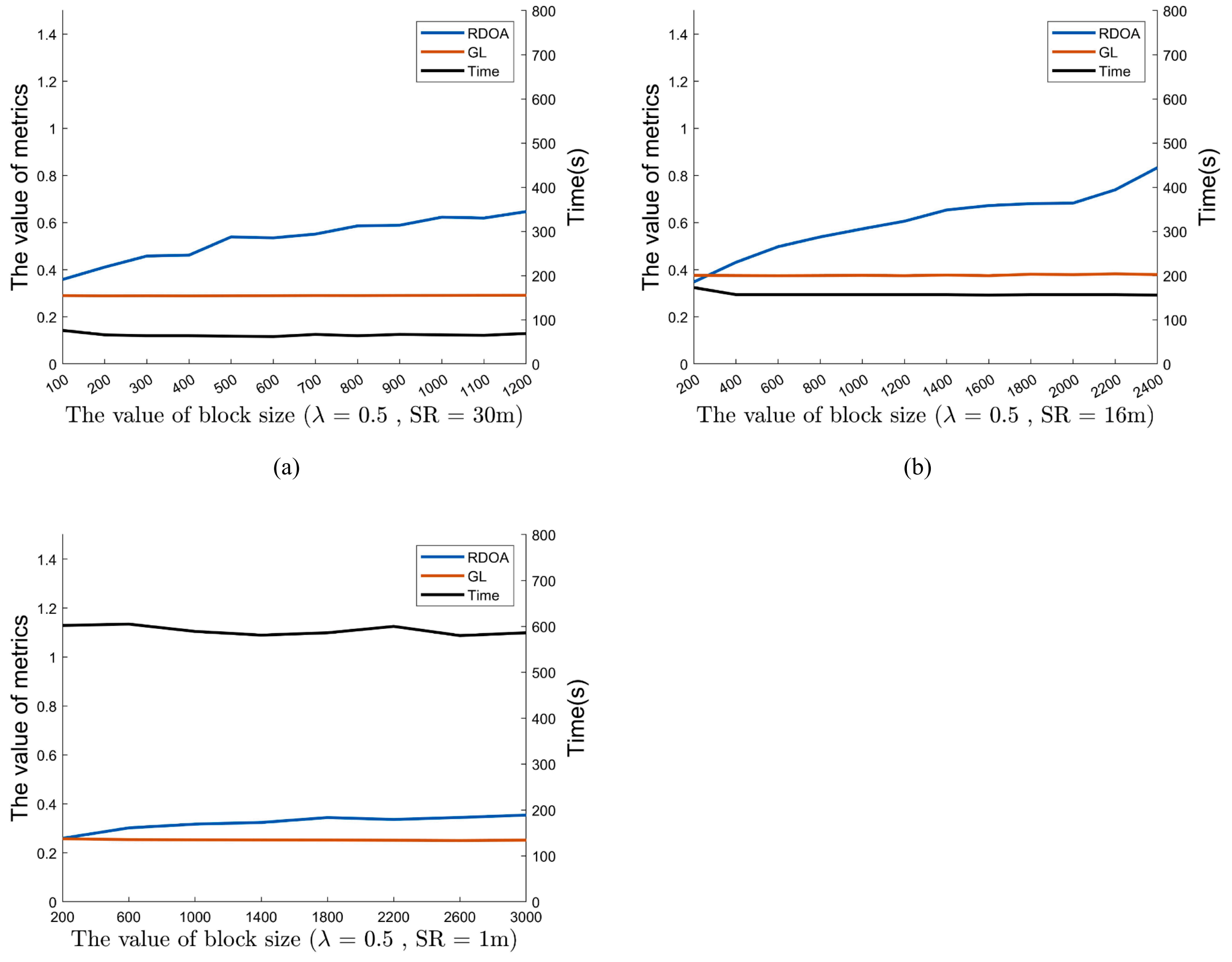
In Fig. 7, the x-axis is the value of λ, the left y-axis is the value of the metric, and the right y-axis is the time in seconds. Fig. 7 (a)–(f) show the variation of *RDOA*, *GL*, and *Time* during the variation of λ from 0.3 to 10 for image block sizes of 100, 200, 300, 400, 600, and 800, respectively. It can be observed that the *RDOA* values all show an increase and then a decrease with the growth of λ for different image block sizes, and then remain stable at some stage. However, *GL* and *Time* are largely unaffected by changes in the value of λ and the image block size and remain at a low value. Therefore, to balance the RN capability with the processing efficiency, a range of λ of [0.5, 1] is appropriate.

In Fig. 8, we show the visual effect of the RN results as λ increases when the image block size is set to 200. As can be seen from the full map, all the results show a good performance in terms of overall radiometric tonal consistency. From the local zoomed-in map, when λ is between [0.5, 1], there are no visible local radiometric differences and the transitions are smooth. However, when λ is greater than 1 and continues to grow, the radiometric differences between images tend to become slightly larger in some transition areas, which is consistent with the trend that can be observed for the *RDOA* value in Fig. 7. It can be seen that when the range of λ is [0.5, 1], better visual and quantitative evaluation results can be obtained. In summary, this parameter is insensitive. Moreover, for all the datasets in this study, we found that the proposed approach has a very good performance when λ is 0.5, i.e., it can effectively complete the RN in less time. Therefore, λ was set to 0.5 in the experiments conducted in this study.

3.3.2. Image block size

Different image block sizes can directly affect the effect of local RN. Since the feature richness of images with different spatial resolutions tends to be different, different image block sizes are often required for images with different spatial resolutions. In this experiment, datasets #1, #3, and #4 with spatial resolutions of 30 m, 16 m, and 1 m were used to quantitatively evaluate the influence of this parameter. The results for dataset #3 are provided for visual evaluation in Fig. 9 and Fig. 10. Note that a square block was used in the experiments.

From Fig. 9, it can be observed that, for datasets with different spatial resolutions, the *RDOA* value becomes larger as the image block size increases, while the *GL* is consistent with the previous experimental



(a)

(b)

(c)

Fig. 9. Illustration of the influence of the image block size. Note that SR represents spatial resolution.

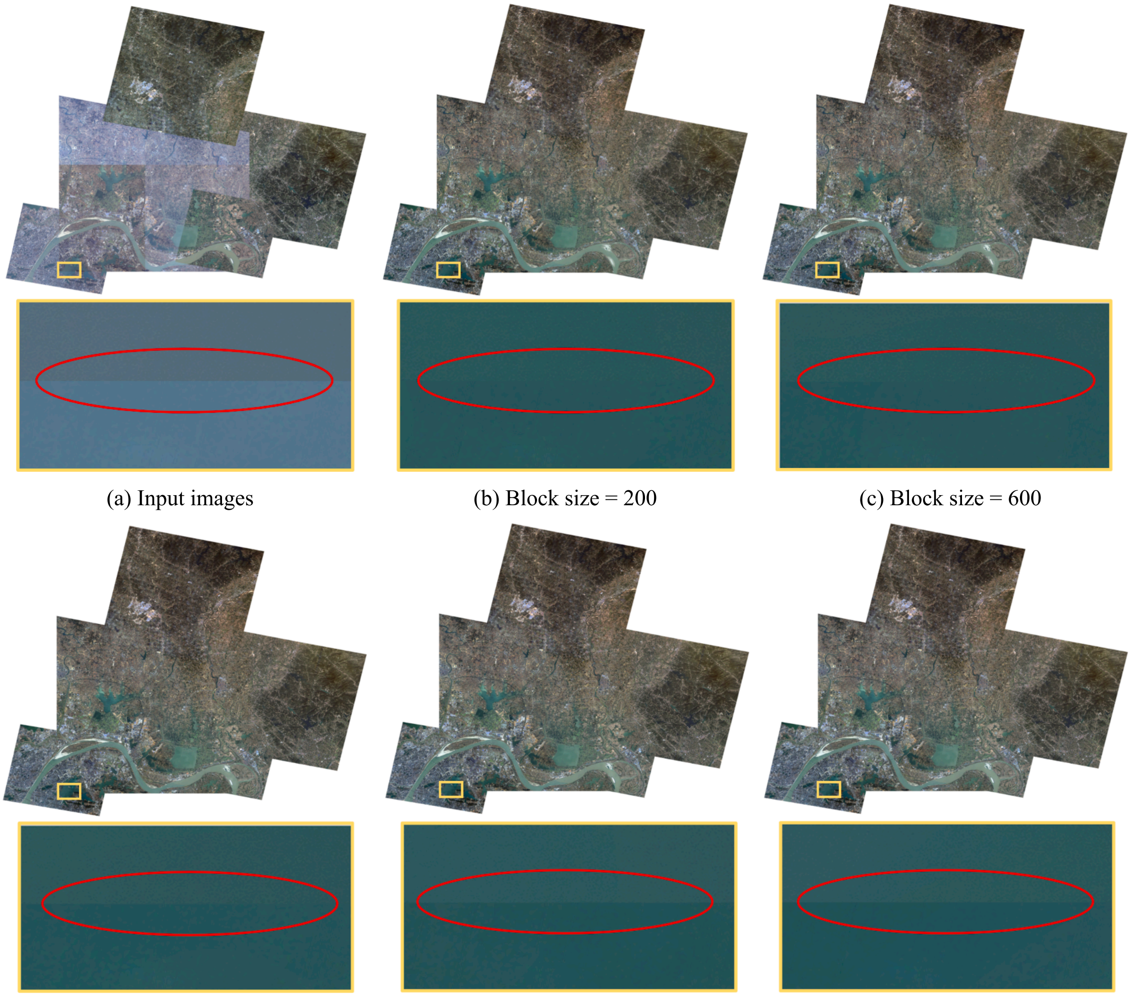
results, and is not affected by the change in the image block size and maintains a lower value. *Time* is less affected by the image block size and tends to decrease slightly as the image block size becomes larger. Considering the effect of the local RN and processing efficiency, we consider that spatial resolutions of 30 m, 16 m, and 1 m with image block sizes of [200, 400], [400, 600], and [600, 1200], respectively, are appropriate.

In Fig. 10, we show the visual effect of the change in the radiometrically normalized images as the image block size increases for dataset #3. When λ is set to 0.5, from the full map, all the results have overall radiometric tonal consistency. From the local zoomed-in map, when the image block size is between [200, 600], there are basically no visible local radiometric differences. However, when the image block size is larger than 600 and keeps increasing, the local radiometric differences between images have a slight tendency to become larger in some transition regions, which is consistent with the trend of the *RDOA* values observed in Fig. 10. In summary, this parameter is insensitive. In the experiments, we set the image block size to 200 for datasets #1 (30 m) and #2 (30 m), 400 for dataset #3 (16 m), and 600 for datasets #4 (1 m) and #5 (1 m).

3.4. Ablation study

The most important aspect of the proposed approach is the VOLRN model in the second stage. To verify the effectiveness of this model, the results of the BAGRN model and the BAGRN-VOLRN approach are shown in Fig. 4–Fig. 6**(d) and (f)**, respectively. As can be observed from both the full maps and local zoomed-in maps, the results processed by the BAGRN model all have local feature tone discontinuities and edge artifacts, whereas for the results processed by the BAGRN-VOLRN approach, all these local radiometric discrepancies are eliminated well, and very good local RN results are obtained.

Meanwhile, the quantitative evaluation results of the BAGRN model and the BAGRN-VOLRN approach are also listed in Table 3. For *ADM*, *ADSD*, and *CD*, the evaluation metrics become smaller after application of the VOLRN model, which means that better RN results are obtained in the image intersection areas. The VOLRN model effectively eliminates the local radiometric differences, and the transitions are smoother and more natural. For *GL*, only a slight increase occurs, and the values remain low. While it can be observed that the gradient loss of the proposed approach is mainly due to the BAGRN model in the first stage, the VOLRN model causes only a very small gradient loss. In terms of *Time*, the time difference between the BAGRN model and the BAGRN-VOLRN approach is not significant. This is because, in the specific programming


(a) Input images (b) Block size = 200 (c) Block size = 600

(d) Block size = 1000 (e) Block size = 1400 (f) Block size = 1800

Fig. 10. Visual illustration of the influence of the image block size for dataset #4.

implementation, some of the image reading and writing tasks of the BAGRN-VOLRN approach can be merged when performing global and local RN, while the model solving is less time-consuming, so the overall processing time difference is smaller, and thus a higher processing efficiency can be achieved.

In addition, the differences between the RN results for datasets #2 processed by the BAGRN model and the BAGRN-VOLRN approach are shown in Fig. 11. It can be observed that the VOLRN model mainly performs further fine RN on the overlapping areas and the intersections of overlapping and non-overlapping areas, while less RN is performed on non-overlapping areas, with a consistent performance in the red, green, and blue bands. This is consistent with the modeling rationale described in Section 2.3.2, in that the model can perform well and maintain sparsity by allowing RN of only certain image blocks. Therefore, if only the BAGRN model is used, it is difficult to eliminate the local radiometric differences and obtain smooth and natural transitions at the image intersections. If only the VOLRN model is used, it is difficult to effectively normalize the overall radiometric differences between images, and it is not possible to achieve a better overall radiometric tonal consistency.

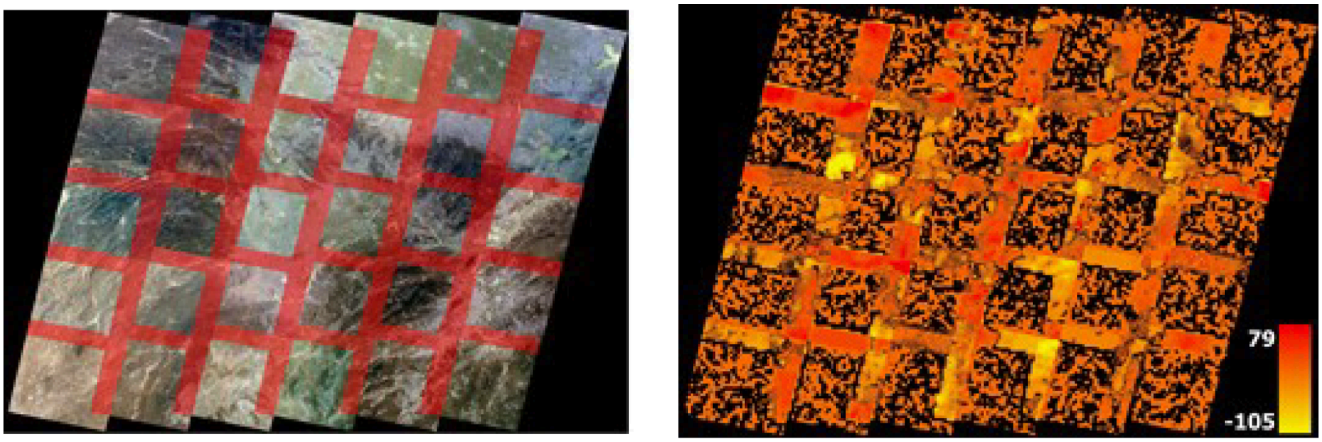
In summary, both stages of the proposed approach are essential. The first-stage model effectively normalizes the overall radiometric tonal differences between images, while the second-stage model effectively eliminates the remaining local radiometric discontinuities and edge artifacts. As a result, the RN results with overall radiometric tonal consistency and smooth and natural local radiometric transitions can be obtained.

4. Conclusion

In this paper, we have proposed a two-stage approach for global and local RN to generate mosaicked images with overall radiometric tonal consistency, natural local radiometric transitions, and invisible seamlines. The key advantages of the proposed approach are that it can eliminate the global and local radiometric differences between multiple images, preserve the image gradients well, and has high computational efficiency. The main contributions of this paper can be summarized as follows:

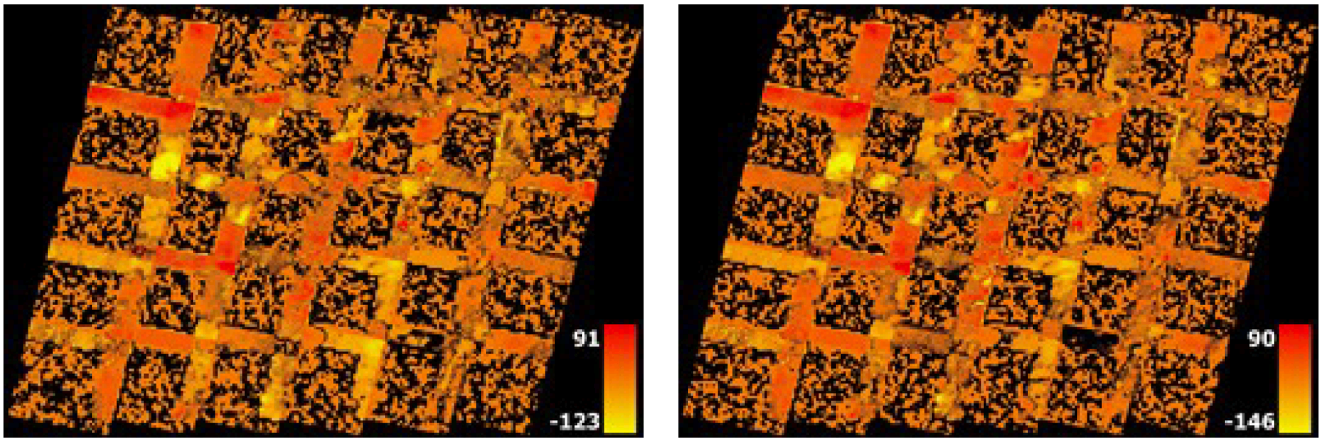
- A novel RN framework that integrates joint block adjustment and variational optimization has been introduced, which effectively addresses global and local radiometric differences. The proposed approach initially balances the overall radiometric tones and subsequently eliminates the residual local radiometric differences, achieving both global and local radiometric consistency and natural transitions in images.
- The VOLRN model utilizes a variational optimization framework with ℓ_1 norm constraint within a unified energy function, allowing flexible adjustment of the local radiometric differences, thereby effectively normalizing the intersection and transition areas in the images.

The comparative experimental results obtained on three challenging datasets of cross-sensor and multi-temporal remote sensing images



(a) Original images

(b) Difference image of red band



(c) Difference image of green band

(d) Difference image of blue band

Fig. 11. Differences between the BAGRN model and the BAGRN-VOLRN approach of datasets #2. (a) is the original images with the overlapping areas labeled. (b) – (d) are the difference images of red, green, and blue bands.

demonstrate that the proposed approach significantly improves both the visual quality and quantitative metrics, outperforming the state-of-the-art RN approaches. Ablation experiments confirmed the indispensability of the two-stage model in the proposed approach, which can synergistically eliminate both the global and local radiometric differences in multiple remote sensing images.

In our future work, more satellite images from different sensors, aerial images and UAV images will be used for further evaluation to improve the reliability and stability of the approach. In addition, since high-resolution remote sensing data often have huge data volume, and the reading and writing of data take up more time in the process of RN, so it is necessary to make sufficient improvements in terms of algorithm optimization and program parallelism to enhance the overall processing performance.

CRediT authorship contribution statement

Dekun Lin: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Huanfeng Shen:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Xinghua Li:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization. **Chao Zeng:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Tao Jiang:** Validation, Software, Formal analysis, Data curation. **Yongming Ma:** Writing – original draft, Methodology, Formal analysis, Data curation. **Mingjie Xu:** Writing – original draft, Validation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 42130108; in part by the Key R&D Project of Hubei Province, China, under Grant 2023BAB066; and in part by the Fundamental Research Funds for the Central Universities under Grant 2042023kfyq04. In addition, the authors would like to thank Dr. Li Li and Dr. Yin Hongche from the School of Remote Sensing and Information Engineering at Wuhan University for their assistance during the experimental process.

References

Boyd, S., 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends® Mach Learn. 3, 1–122. https://doi.org/10.1561/2200000016.

Brown, M., Lowe, D.G., 2007. Automatic panoramic image stitching using invariant features. Int. J.Comput. vis. 74, 59–73. https://doi.org/10.1007/s11263-006-0002-3.

Chen, X., Vierling, L., Deering, D., 2005. A simple and effective radiometric correction method to improve landscape change detection across sensors and across time. Remote Sens. Environ. 98, 63–79. https://doi.org/10.1016/j.rse.2005.05.021.

Cresson, R., Saint-Geours, N., 2015. Natural color satellite image mosaicking using quadratic programming in decorrelated color space. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4151–4162. https://doi.org/10.1109/JSTARS.2015.2449233.

Hong, Z., Xu, C., Tong, X., Liu, S., Zhou, R., Pan, H., Zhang, Y., Han, Y., Wang, J., Yang, S., 2023. Efficient global color, luminance, and contrast consistency optimization for multiple remote sensing images. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 16, 622–637. https://doi.org/10.1109/JSTARS.2022.3229392.

Li, X., Feng, R., Guan, X., Shen, H., Zhang, L., 2019b. Remote sensing image mosaicking: achievements and challenges. IEEE Geosci. Remote Sens. Mag. 7, 8–22. https://doi.org/10.1109/MGRS.2019.2921780.

Li, X., Hui, N., Shen, H., Fu, Y., Zhang, L., 2015. A robust mosaicking procedure for high spatial resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 109, 108–125.

Li, X., Li, Z., Feng, R., Luo, S., Zhang, C., Jiang, M., Shen, H., 2020b. Generating high-quality and high-resolution seamless satellite imagery for large-scale urban regions. Remote Sens 12.

Li, S., Wang, H., Wang, L., Yu, X., Yang, L., 2019a. Parallel color balancing method using adaptive block Wallis algorithm for image mosaicking. J Remote Sens 23, 706–716.

Li, L., Xia, M., Liu, C., Li, L., Wang, H., Yao, J., 2020a. Jointly optimizing global and local color consistency for multiple image mosaicking. ISPRS J. Photogramm. Remote Sens. 170, 45–56. https://doi.org/10.1016/j.isprsjprs.2020.10.006.

Li, Y., Yin, H., Yao, J., Wang, H., Li, L., 2022. A unified probabilistic framework of robust and efficient color consistency correction for multiple images. ISPRS J. Photogramm. Remote Sens. 190, 1–24. https://doi.org/10.1016/j.isprsjprs.2022.05.009.

Lin, D., Shen, H., Qiu, Z., Zhu, S., Huang, W., Jiang, T., 2024. A framework for generating high-resolution seamless remote sensing images for regional-scale areas. IEEE Geosci. Remote Sens. Lett. 21, 1–5. https://doi.org/10.1109/LGRS.2024.3380165.

Liu, K., Ke, T., Tao, P., He, J., Xi, K., Yang, K., 2020. Robust Radiometric Normalization of Multitemporal Satellite Images Via Block Adjustment Without Master Images. IEEE J. Sel. Top. Appl. EARTH Obs. REMOTE Sens. 13.

Oliveira, M., Sappa, A.D., Santos, V., 2015. A Probabilistic Approach for Color Correction in Image Mosaicking Applications. IEEE Trans. Image Process. 24, 508–523. https://doi.org/10.1109/TIP.2014.2375642.

Pan, J., Wang, M., Li, D., Li, J., 2010. A network-based radiometric equalization approach for digital aerial orthoimages. IEEE Geosci. Remote Sens. Lett. 7, 401–405.

Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P., 2001. Color transfer between images. IEEE Comput. Graph. Appl. 21, 34–41. https://doi.org/10.1109/38.946629.

Shen, T., Wang, J., Fang, T., Zhu, S., Quan, L., 2016. Color Correction for Image-Based Modeling in the Large, in: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (Eds.), Computer Vision – ACCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 392–407. Doi: 10.1007/978-3-319-54190-7_24.

Sui, H., Liu, C., Liu, J., Zheng, X., Li, H., Yu, S., Li, Q., 2020. Reflection and exploration of rapid remote sensing emergency response for typical natural disasters. Geomat. Inf. Sci. Wuhan Univ. 45, 1137–1145.

Yu-Wing Tai, Jiaya Jia, Chi-Keung Tang, 2005. Local Color Transfer via Probabilistic Segmentation by Expectation-Maximization, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE, San Diego, CA, USA, pp. 747–754. Doi: 10.1109/CVPR.2005.215.

West, H., Quinn, N., Horswell, M., 2019. Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities. Remote Sens. Environ. 232, 111291 https://doi.org/10.1016/j.rse.2019.111291.

Xia, M., Yao, J., Xie, R., Zhang, M., Xiao, J., 2017b. Color Consistency Correction Based on Remapping Optimization for Image Stitching, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Presented at the 2017 IEEE International Conference on Computer Vision Workshop (ICCVW), IEEE, Venice, pp. 2977–2984. Doi: 10.1109/ICCVW.2017.351.

Xia, M., Yao, J., Xie, R., Li, L., Zhang, W., 2017a. Globally consistent alignment for planar mosaicking via topology analysis. Pattern Recognit. 66, 239–252. https://doi.org/10.1016/j.patcog.2017.01.020.

Xia, M., Yao, J., Gao, Z., 2019. A closed-form solution for multi-view color correction with gradient preservation. ISPRS J. Photogramm. Remote Sens. 157, 188–200. https://doi.org/10.1016/j.isprsjprs.2019.09.004.

Xiang, Y., Zou, B., Li, H., 2009. Selective color transfer with multi-source images. Pattern Recognit. Lett. 30, 682–689. https://doi.org/10.1016/j.patrec.2009.01.004.

Xiong, Y., Pulli, K., 2010. Color matching of image sequences with combined gamma and linear corrections. In: Proc. Int. Conf. ACM Multimedia. pp. 261–270.

Yu, L., Zhang, Y., Sun, M., Zhou, X., Liu, C., 2017. An auto-adapting global-to-local color balancing method for optical imagery mosaic. ISPRS J. Photogramm. Remote Sens. 132, 1–19. https://doi.org/10.1016/j.isprsjprs.2017.08.002.

Zhang, X., Feng, R., Li, X., Shen, H., Yuan, Z., 2022. Block adjustment-based radiometric normalization by considering global and local differences. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2020.3031398.

Zhou, X., 2015. Multiple auto-adapting color balancing for large number of images. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-7/W3, 735–742. https://doi.org/10.5194/isprsarchives-XL-7-W3-735-2015.