



# Differential evolution algorithm with population knowledge fusion strategy for image registration

Yu Sun<sup>1</sup> · Yaoshen Li<sup>1</sup> · Yingying Yang<sup>1</sup> · Hongda Yue<sup>1</sup>

Received: 31 October 2020 / Accepted: 12 April 2021 / Published online: 3 May 2021  
© The Author(s) 2021

## Abstract

Image registration is a challenging NP-hard problem within the computer vision field. The differential evolutionary algorithm is a simple and efficient method to find the best among all the possible common parts of images. To improve the efficiency and accuracy of the registration, a knowledge-fusion-based differential evolution algorithm is proposed, which combines segmentation, gradient descent method, and hybrid selection strategy to enhance the exploration ability in the early stage and the exploitation ability in the later stage. The proposed algorithms have been implemented and tested with CEC2013 benchmark and real image data. The experimental results show that the proposed algorithm is superior to the existing algorithms in terms of solution quality, convergence speed, and solution success rate.

**Keywords** Remote sensing image · Differential evolution · Image registration · Knowledge fusion

## Introduction

Image registration is a complex task in image processing, which refers to match different images of the same scene taken at multiple times, in multiple viewpoints or with multiple sensors [21,22]. Remote sensing image registration methods proposed in literature consist of two categories: feature-based registration methods and intensity-based registration methods. The feature-based registration method extracts the prominent features of an image, which is considered as control points. In general, the point is the shape contours represented by edge points, or point features represented by the positions of interest points. The registration problem is to find the correct correspondence between two sets of points extracted from the input data and restore the

underlying spatial mapping at the same time, which can also be considered a point set registration problem. The widely used solutions, such as scale invariant feature transform (SIFT) [19,20], speeded up robust features (SURF) [2], and shape context (SC) [3], are still hot spots in image registration. The second is intensity-based registration method [7,27]. In this method, the intensity of images is used as a measure of similarity between images. Intensity-based method usually contains two steplike components: similarity measurement and algorithm optimization. Similarity measurement is a key step in intensity-based registration method and appropriate similarity measurement directly affects the registration results. Many similarity measurements have been proposed, such as sum of squared differences (SSD) [9], cross correlation (CC) [44], normalization mutual information (NMI) [4].

The evolutionary algorithm is easy to find the best corresponding point or transformation parameter from image registration problem, which derived many other important algorithms such as genetic algorithm (GA) [1], particle swarm optimization (PSO) [11], and differential evolution algorithm (DE) [6]. Evolutional algorithm is able to solve various problems with improved strategies like mixed-variable problem [37] and nonlinear programming [35], and dynamic multi-objective optimization problems [36] with improved Prediction strategies. Several improved algorithms were proposed to get the best multi-modal problems. Qu et al. [25]

✉ Yu Sun  
sunyu@gxu.edu.cn

✉ Yaoshen Li  
a17687508651@163.com

Yingying Yang  
berlinyang1034@163.com

Hongda Yue  
yuehongda1017@163.com

<sup>1</sup> School of Computer and Electronics and Information, Guangxi University, Nanning 530004, People's Republic of China

put forward a neighborhood mutation strategy which is also called niche technology, and proposed CDE [29], SDE [18], NCDE [25], NSDE [25] etc. by combining them with DE to solve multi-modal problems. ADE and SaEPSDE were proposed by Lacca et al. [10] besides NSHDE [25] are a multi-strategy DE, which improves the population diversity by using multiple evolution strategies. Li et al. [13] proposed R2PSO, R3PSO, R2PSOLHC, R3PSOLHC and Ye et al. [41] proposed MO-Ring-PSO-SCD, and all these algorithms used the PSO algorithm with ring topology to divide subpopulation, maintaining the population diversity. Wang et al. [34] proposed an enhance DE with niching technique and adaptive learning strategy can be used to enhance the exploration ability of the algorithm.

To seek and locate multiple optimal solutions, niche technology [24,31] is used in the multi-modal optimization problems. Niche technology is to divide the population of each generation into several sub-populations, the crossover and mutation in the sub-populations or among different sub-populations produce a new generation of populations. At the same time, the selection strategies of pre-selection, exclusion, and sharing are used to select individuals that will be retained to the next generation. The evolutionary algorithm combined with niche technology can maintain population diversity and has a high global optimization capability and convergence speed. Hence, it is especially suitable for solving MMOP problems. In the last decades, lots of scholars have proposed many technologies about niches, including crowding [32], speciation [12], fitness sharing [5], valleys [33], and clustering [25,42] etc. The main process is to divide the entire population into several niches and seek optima in each niche. For the simple and effective of DE, several improved DE combined niche technology were proposed in recent years. Wang et al. [39] proposed a niche DE based on the minimum spanning tree (MSTDE). In this method, the MST in each iteration is constructed based on the distance information between individuals, and the population is divided into several niches based on the  $M$  maximal weighted edges of MST. Besides, a dynamic pruning ratio (DPR) strategy is used to determine the size of the niche  $M$  to improve the performance of the algorithm. Experiments show that the performance of MSTDE is better than other state-of-art multi-modal optimization algorithms when evaluated on the benchmark test functions from CEC2013. Poláková et al. [23] proposed an adaptive method of population size during population evolution in the same year. This method dynamically reduces or increases the size of the population during the evolution process by detecting changes in the diversity of the population. Experimental results show that DE with this new self-adaptive variant has greatly improved the efficiency of the algorithm and is proved effective in optimizing more complex problems such as multi-modal, mixed, or combined problems. Zhao et al. [45] proposed a DE algorithm based

on local binary pattern(LBPADE). LBP can extract relevant information among individuals, and find multiple optima in MMOP. Zhao et al. [45] proposes an adaptive algorithm based on LBP, which uses local binary operators to find multiple optima and divides niches based on these optima. In addition, to improve the exploration and exploitation capabilities of the algorithm, the mutation strategy and parameter strategy of the algorithm are improved, hence, niche and global interaction (NGI) mutation strategy and adaptive parameter strategy (APS) are proposed. The NGI mutation strategy uses the information of the niche and the global space to further explore the current search space, while the APS adjusts the parameters according to the LBP information of individuals to let individuals close to the optima. Results on the MMOP test functions show that the performance of LBPADE is superior to most state-of-art algorithms.

To improve the efficiency, some novel methods are employed for registration. Knowledge fusion, developed from information fusion, is a process where knowledge from different sources interacts to form new knowledge [28,43]. It is commonly used in many engineering areas [26]. It contains the process of abstracting, summarizing, and classifying real-world information and raising them to the aspect of knowledge. Knowledge fusion is applied to enhanced the differential evolution algorithm to get the best solution [38].

To improve the performance of the DE, a novel algorithm, based on niche technology, three different strategies are designed according to three knowledge of the population, are proposed. Besides, a multi-modal DE algorithm (SGD-DE) with a species gradient descent method is proposed. To evaluate the performance of our proposed algorithm, UCAS-ADO dataset and the remote sensing image of Guangxi University as registration image data. Experimental results show that the algorithm proposed in this paper can accelerate the convergence speed of DE and has better performance than traditional DE and other multi-modal optimization algorithms in terms of solution quality and success rate.

The rest of this paper is organized as follows. In Related work section, image registration and optimization techniques are discussed. In Knowledge fusion based DE section, the proposed Knowledge fusion based DE for image registration is presented. Experimental results are described in Experiment section. Finally, conclusions are drawn in Conclusion section.

## Related work

### Image registration problem

Image registration refers to register two or more images collected at different times, in different spaces, and with different equipments into a clear image with objects all focused, which is denoted as



**Fig. 1** Left is Guangxi University in 2011, and right is Guangxi University in 2013

$$T = \underset{T}{\operatorname{argmin}}(E(R - T \otimes S)), \quad (1)$$

where  $T$  is the affine transformation parameter of the image,  $T \otimes S$  represents the affine transformation operation performed on the image  $S$ ,  $T$  is the affine transformation matrix.  $S$  is the image to be registered, and  $R$  is the sample image.  $E(R - T \otimes S)$  represents the similarity measurement function of the two images. According to different functions, it can be divided into two cases: maximum problem and minimum problem. For example, the higher the image similarity, the larger the function value of MI and NMI [8], but DTV [16] is the opposite. The purpose of image registration is to find the transformation parameter that optimizes the similarity measurement function between images. DTV [16] is an optimal similarity measurement method based on the gradient domain [17] of total variation. DTV is designed as a similarity measure to match the edges of two images. When the edge features of the two images correspond, DTV can be expressed by calculating the image gradients. The size and position of the image gradient should be similar, and the gradient of the residual image formed after registration should be more sparse because any registration error may produce ghost images and increase the sparsity of the residual image. The DTV function used for image registration is expressed as:

$$\min_T(E(T)) = \|\nabla R - \nabla S(T)\|, \quad (2)$$

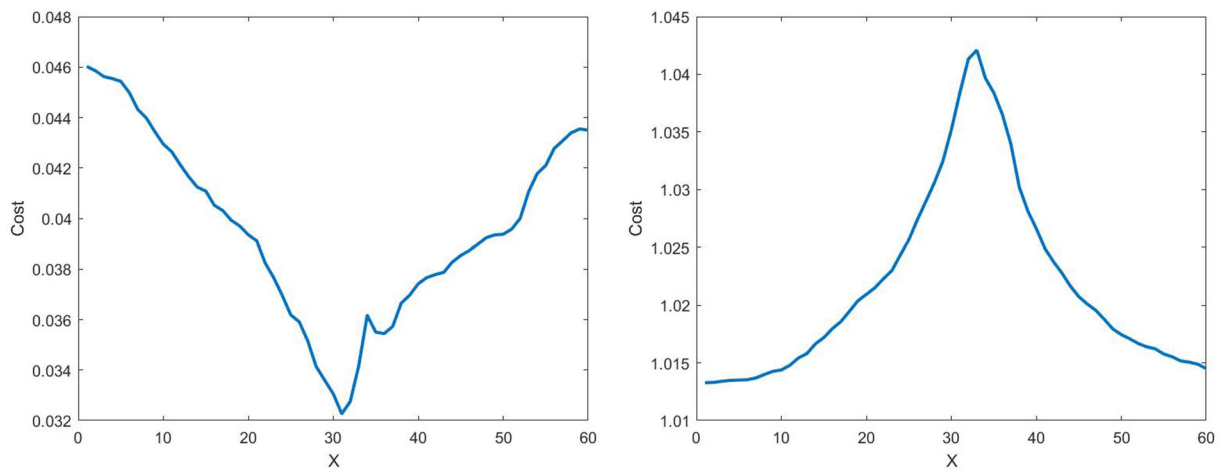
where  $\nabla R = \sqrt{(\nabla_x R)^2 + (\nabla_y R)^2}$  represents the image gradient in the two spatial directions,  $\nabla_x R$  and  $\nabla_y R$  represents the positive finite-difference operator of  $x$  and  $y$  coordinates. DTV is one of the most advanced methods for processing

image registration, it has lower computational complexity and is more accurate and robust than MI and NMI. Figure 2 shows two images at the same location and different time in Fig. 1. After slightly changing the transformed parameters, it can be seen that the fluctuation of DTV is greater than the fluctuation of NMI and MI. That is to say, DTV behaves more sensitive than the other two similarity functions for slight changes in transformed parameters.

From Fig. 2, it can be seen that one-dimensional DTV functions have multiple local optima, but the dimension in practical application is greater than or equal to three. Therefore, image registration is a multi-modal problem. The differential evolution algorithm based on niche technology solve the problem efficiently.

## Differential evolution algorithm

DE [30,40] is a population-based algorithm for optimization problems proposed in 1997 by Storn and Price, DE has relatively high search accuracy, robustness, and good convergence speed. It can be used to find the optimal solutions of nonlinear, non-differentiable, and multi-modal continuous space functions with real-valued parameters. It has the same structure and steps as classical evolutionary algorithms, such as population initialization, crossover, mutation, and selection, however, when generating new candidate solutions, DE uses differential evolution operators. For each generation, operators continue working until the pre-defined terminating condition is met. DE algorithm needs to define control parameters (i.e., population size NP, scaling factor  $F$ , and crossover probability CR, searching space  $\Omega$  In this



**Fig. 2** The changes of the DTV value and the NMI value according to different rotation angles of the two overlapping pictures in Fig. 1

paper, we suppose that the object function to be minimized is  $F(X_i)$ . Firstly, a feasible searching space  $\Omega$  is defined,  $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,D}\}$  is randomly generated in  $\Omega$ ,  $X^U$  and  $X^L$  are the upper and lower bounds of searching space.

**Population initialization** The initial population consists of NP decision vectors (individuals) that are generated by assigning values to each component of all vectors. An optimization problem with D-dimension can be denoted as a vector with D-dimension. DE is based on the differences of individuals and each individual weights can be determined as:

$$X_{i,j} = X_j^U + \text{rand}(0, 1) * (X_j^U - X_j^L), \quad (3)$$

$X_1, X_2, \dots, X_{NP}$  are generated individuals, where each  $X_i$  consists of D vectors,  $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,D}\}$ . Mutation operation: After initialization, DE performs mutation operation for each individual  $X_i$ . Each individual has a correspond mutation individual  $V_i$ . In this paper, DE/rand/1 is conducted to generate mutant vectors. Each mutation individual's weight can be determined as:

$$V_i = X_{r1} + F * (X_{r2} - X_{r3}), \quad (4)$$

where  $V_i$  is the mutant individual,  $X_{r1}, X_{r2}$  and  $X_{r3}$  are three mutually different individuals which are randomly selected from the whole population,  $r1, r2$  and  $r3$  are random integers in the range of  $\{1, 2, \dots, NP\}$  and should be different from the running index  $i$ . Hence, the number of population or niche should be greater than 3.  $F$  is the scaling factor that lies in the range of  $[0, 1]$  for scaling the difference vectors. If the  $F$  value is lower, the convergence speed is faster, while the larger the value, the greater the population diversity.

**Crossover operation** To increase population diversity, DE utilizes crossover operation to integrate mutant individuals and successful individuals reserved from the last generation, trial vectors are selected from mutant vectors and target vectors according to the following formula:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } \text{rand}(0, 1) \leq CR \\ x_{i,j}, & \text{otherwise} \end{cases} \quad (5)$$

Where  $CR$  is the crossover probability, which is usually within the range of  $[0, 1]$ , and  $u_{i,j}$  can be determined whether to assign the mutant vector  $v_{i,j}$  to the trial vector  $u_{i,j}$  by the comparison result between the  $CR$  and a random number  $j$  generated from the range of  $[0, 1]$ .

**Selection operation:** The selection process is the simple competition between offspring and corresponding parents. To confirm whether offspring individuals are reserved in the next generation, the greedy criterion is used to make the comparison, and those who have better fitness values are retained to the next generation. If and only if, the trial vector  $U_{i,G+1}$  yields a fitness value no more than the target vector  $X_{i,G}$ ,  $U_{i,G+1}$  is set to replace  $X_{i,G+1}$ , otherwise  $X_{i,G}$  is kept to the next generation.

$$X_{i,G+1} = \begin{cases} U_{i,G+1}, & \text{if } f(U_{i,G+1}) \leq f(X_{i,G}) \\ X_{i,G}, & \text{otherwise} \end{cases} \quad i = 1, 2, \dots, NP \quad (6)$$

## Knowledge fusion based DE

In this section, the proposed multi-modal knowledge fusion based algorithm for image registration is described in detail. First, the Knowledge representation based population model is introduced, and then the four main steplike procedures is described. The DTV function is used as the object function to find the best image registration solution. The proposed method achieve a good balance between exploration and exploitation through techniques such as niche technology, Gradient descent, multiple operators selection scheme and self-adaptive population updating strategies.



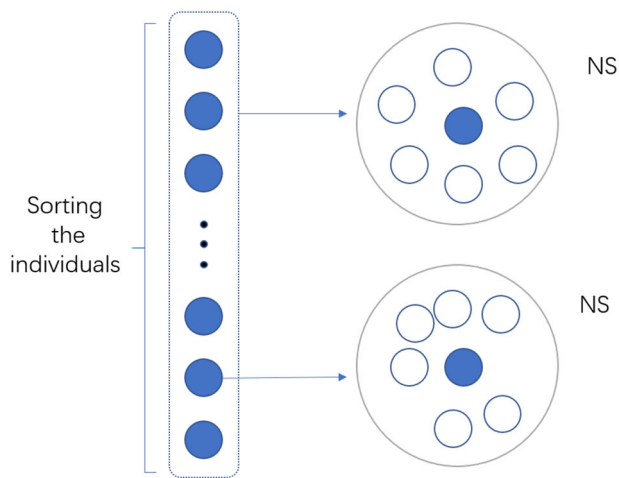


Fig. 3 Division of niches

### Knowledge represent population

In order to make full use of the fitness knowledge of DTV function, niche technology is applied to divide the initialized population into several niches. Each niche searches for its own optimal value to avoid the entire population falling into local optima. Firstly, the proposed algorithm uses the clustering framework of species formation in the niche dividing stage after initialization and combines three different types of knowledge to fuse with other information. Based on that, three new strategies are designed. In the mutation stage, traditional DE operators and the transition probability are utilized, and the state of the neighboring niche is regarded as the first knowledge to guide the fusion of information among niches; In the selection stage, the rank of the best individual’s fitness value of every niche in the entire population is regarded as the second knowledge to select the appropriate selection strategy; In the stage of adjusting the parameters of niches, the distances of central value among niches are regarded as third knowledge to adjust the size of each niche.

### Initialization with niche technology

After population initialization, the clustering method is used to divide the niches. This paper uses the clustering framework of species formation as the method of dividing the niche. NP is the population size, NS is the initial size of the niche, and each niche constructs NS new solutions in each iteration. The Fig. 3 shows the division of niches.

The first step is to sort the current population according to the fitness value, and the individual with the best fitness value is set as the seed of the first niche; the second step is to select NS-1 individuals closest to the seed to form a new niche; Finally, remove these NS individuals from the popu-

lation, and repeat the above three steps until no remaining individuals in the population.

### Mutation with gradient descent

After initializing the population and dividing the niches, ideally, the entire population can roughly cover the entire solution space, and all the niches are evenly distributed in the entire solution space. In every iteration, each niche evolves independently. However, not every niche has local optima. If the niche is closed to evolve, it is not conducive to the evolution of the entire population. Therefore, a mutation strategy is used to promote communications among niches, which is to say, the state of the neighboring niche is regarded as the first knowledge to guide the fusion of information among niches. Firstly, a static transition probability PC is defined.

Start looping from the first individual, and when a random number generated from the range of [0, 1] is larger than PC, that is  $rand > PC$ , the current individual uses a search strategy with gradient descent in the mutation stage

$$dis_m = \lambda * (X_{best-nearest}(t) - X_i) \tag{7}$$

$$V_m = X_i + dis_m, \tag{8}$$

dis is the step distance,  $\lambda$  is set as 0.1, i is the loop number of gradient descent and  $V_i$  is the intermediate vector produced in the  $i$ th descend.  $X_{best-nearest}$  is the optimal solution from the nearest niche that has better optimal fitness than the current niche. The replacement strategy is used to update the current individual in every iteration:

$$X_i = \begin{cases} V_i, & \text{if } f(V_i) \leq f(X_i) \\ X_i, & \text{otherwise} \end{cases} \tag{9}$$

When  $f(f(V_i) > f(X_i))$ , let  $\lambda = \alpha * \lambda$ ,  $\alpha$  is a value less than 1. Increase i in every iteration until it reaches the maximum iteration. When the random number is less than PC, that is  $rand(0, 1) < PC$ , the current individual is denoted as follows:

$$v_i = x_{r1,i} + F * (x_{r2,i} - x_{r3,i}), \tag{10}$$

where  $r1, r2, r3$  are random numbers from 1 to NS and  $F$  is the scaling factor. An intermediate vector is generated following the same formula as the standard DE algorithm. The selection of the intermediate vector also follows the same process as the standard DE algorithm, and new content is added. Details are presented in the following sections.

Through this mutation strategy, individuals in the niche are more likely to be close to the optima of the nearest niche.

### Dual-selection method

After dividing the niches, evolutionary operations are performed in each niche to generate offspring. On the basis of that, selection operations are performed to chose individuals that can be retained to the next generation, and form

new niche groups. So far, there are mainly two selection operators, and both have been widely used in multi-modal algorithms. One is combination selection that first combines NP parents with NP offspring (NP is the population size), and then selects the best N individuals from 2NP individuals. Another is the one-by-one selection which compares the fitness value of each offspring with its nearest parent individual, generally using Euclidean distance to measure. If the offspring has better fitness value, then replace the parent with the corresponding individual, which is to say, the rank of the best individual's fitness value of every niche in the entire population is regarded as the second knowledge to select the appropriate selection strategy;

The two selection operators are respectively beneficial to the evolution of the entire population. The one-by-one selection operator, which selects the parent with the most similar genes to the offspring, and then compares these two individuals and replaces the poorer one with the best one, which can maintain the population diversity and enhance the exploration capabilities; while the combination selection operators, mixing offspring and parent individuals, select the best NP individual to enter the next generation, which can fasten the convergence speed of the population, further improve the accuracy of the optimal solution, and improve the exploitation capabilities of the algorithm. Therefore, the strategy proposed in this paper is to select different selection operators according to the fitness value of the current niche.

Firstly, each niche is divided into two parts according to the fitness value of the optimal individual. The first part, called the superior niche, consists of the niches where the current optimal solutions of the entire population is located, and the second part, called the inferior niche, consists of the rest niches. The goals of the two parts of the niche are different, and the replacement strategy is selected based on this.

Secondly, If the current individual belongs to the superior niche, such as individual a in niche A, it should further explore its neighboring individuals to accelerate convergence speed, and use the combination selection strategy for the offspring of this niche: select N best individuals from 2N individuals (N parent individuals, N offspring individuals) to improve the accuracy of the solution. Conversely, if the current individual belongs to inferior niche, such as individual b in niche B, use a one-by-one selection operator for its evolved offspring: by comparing the fitness value of offspring individual and parent individual with the closest Euclidean distance from the offspring, and replace the parent individual that is inferior to the offspring, which enables the niche to further explore optima in the search space and maintain the population diversity.

The selection strategy is described in Algorithm 1. If the optimal individual in the niche is equal to the global optimal individual, the combination selection operator is used, the

parent and the offspring are mixed, and the optimal NS individuals are selected after sorting; otherwise, the one-by-one selection is used, comparing the offspring with the parent individual which is the closest to the offspring according to the Euclidean distance and select the better individual to be retained to the next generation.

---

#### Algorithm 1 :Selection strategy

---

**Require:** Niches,Population:NP,Niche Size:NS,gbestval

**Ensure:** Current niche:pop

```

1: for  $j = 1, 2, \dots, NP/NS$  do
2:   compute the fitness  $Cost_j$  of the individual in the niche  $j$ 
3:   if  $\max(Cost_j)=gbestval$  then
4:     newpop = pop  $\cup$   $u_j$ 
5:     newpop = sort(newpop)
6:     pop = newpop[1:NS]
7:   else
8:     for  $k = 1, 2, \dots, NS$  do
9:       popx is the closest individual to  $u_{j,k}$  in this niche  $j$ 
10:      if  $f(u_{j,k}) > f(pop_x)$  then
11:        popx =  $u_{j,k}$ 
12:      end if
13:    end for
14:  end if
15: end for

```

---

#### Self-adaptive updation

To maintain the diversity of each niche, the merge operation is conducted between two niches that are too close to each other. For the  $g$ -th niche, where the individuals are  $x_1^g, x_2^g, \dots, x_{NS}^g$ , firstly calculate the center point  $C^g$  of the NS individuals in each niche:

$$c_j^g = \frac{\sum_{i=1}^{NS} x_{i,j}^g}{NS}, \quad (11)$$

where  $j = 1, 2, \dots, D$ ,  $c_j^g$  represents the center point vector of the  $j$ th variable.

Then, the Euclidean distances among center points of each niche are denoted as follows:

$$d_{g,g'} = \sqrt{\sum_{j=1}^D (c_j^g - c_j^{g'})^2} \quad (12)$$

When  $d_{g,g'} < \tau$ , If the distance between the niches are too close, the two niches will be merged. Because another evolution strategy is used in the mutation step, in this strategy, the niche with poor fitness will move closer to the nearest neighbor with high fitness. Hence, when the distance between the two niches is small enough, merge the two niches, which can ensure the population diversity of each niche. If the current niche falls into a local optimum, the size of the cluster is more likely to increase so that more accurate solutions can be obtained or avoid getting trapped into the local optima.

## The overall procedure of SGD-DE algorithm

The steps of the SGD-DE algorithm are as follows, Algorithm 2 is the pseudocode of the algorithm. The first step is to initialize the population randomly; the second step is to divide the niches according to the principle of species formation; the third step is to evolution operation, individuals are randomly selected based on probability, part of individuals are generated from the local niche, and another part of individuals generate offspring of the next generation according to the standard DE algorithm; the fourth step: selection operation.

---

### Algorithm 2 :SGD-DE

---

**Require:** Population:NP;Dimension:D;Generation:MaxFES;F;CR;Niche Size:NS;Pc

**Ensure:** The best vector(solution)- $\Delta$

```

1: initialize:Current population:pop,gbestval
2: initialize niches according to Fig. 3
3: for  $it = 1, 2, \dots, MaxFES$  do
4:   for  $i = 1, 2, \dots, NP/NS$  do
5:     for  $j = 1, 2, \dots, NS$  do
6:       if  $rand > Pc$  then
7:         Compute  $u_i$  use equation (7)-(9)
8:       else
9:         Compute  $u_i$  according to equation (10)
10:      end if
11:    end for
12:    for  $j = 1, 2, \dots, NP/NS$  do
13:      compute the fitness  $Cost_j$  of the individual in the niche  $j$ 
14:      if  $\max(Cost_j) = gbestval$  then
15:         $newpop = pop \cup u_j$ 
16:         $newpop = \text{sort}(newpop)$ 
17:         $pop = newpop[1:NS]$ 
18:      else
19:        for  $k = 1, 2, \dots, NS$  do
20:           $pop_x$  is the closest individual to  $u_{j,k}$  in this niche  $j$ 
21:          if  $f(u_{j,k}) > f(pop_x)$  then
22:             $pop_x = u_{j,k}$ 
23:          end if
24:        end for
25:      end if
26:    end for
27:  end for
28: end for

```

---

## Experiment

### Parameter settings

In this experiment for the proposed algorithm, we use the fixed parameters NP (population size) = 200,  $F$  (scaling factor) = 0.5, CR (crossover rate) = 0.9, and the initial niche size is set to 10, besides, the iterations are no more than 150. The image datasets in this paper are collected from the UCAS-AOD Dataset and the remote sensing images of Guangxi

University. The CEC2013 multi-modal benchmark functions [15] which contain 12 multi-modal test functions are used to test the proposed algorithm.

### Evaluation index

Peak ratio (PR), representing the average percentage of global peaks found in multiple runs, is used as a evaluation index in this paper and is denoted as follows:

$$PR = \frac{\sum_{i=1}^{NR} NPF_i}{TNP * NR}, \quad (13)$$

where NR is the number of runs,  $NPF_i$  is the number of global peaks found in the  $i$ th run, and TNP is the total number of global peaks in the optimization problem.

### Comparisons with state-of-the-art multimodal algorithms

#### Comparison at accuracy

In order to test the multi-modal performance of the two algorithms proposed in this paper, the CEC2013 multi-modal benchmark function is used to evaluate the performance of SGD-DE in solving MMOPs.

In this section, two experiments are conducted. The algorithms involved in the experiment are CDE, SDE, NCDE, NSDE, NSHDE, and the particle swarm optimization algorithm with ring topology: R2PSO, R3PSO, R2PSOLHC, R3PSOLHC, FERPSO [13]. The total number of individuals for all algorithms is set to 200. The other parameters of the algorithm used for comparison are based on the corresponding references.

CEC2013 multi-modal functions are used to test the multi-modal capabilities of the SGD-DE algorithm, DE number multimodal algorithm, PSO with ring topology, and the mainstream multi-strategy adaptive DE algorithm. The experimental results show in Table 1. In the table, if the PR value is higher than other algorithms or equal to the other algorithms, the result will be highlighted in bold, and the total times of the best results got in each algorithm are counted at the endline of the table.

As shown in Table 1, the best results of SGD-DE were obtained on F1–F5 and F11 when  $\varepsilon = 1.0e-01$ , the results on F6, F8 and F10 are not the best but very close to the best; the results on the remaining functions differ from the best results. Tables 2 and 3 represent the search results at the accuracy of  $1e-2$  and  $1e-5$ , and SGD-DE has the best overall optimization results on F1–F5, F11, F13–F20; the results on F6 and F10 are very close to the best, but the results are different in the remaining functions.

When the accuracy is higher, higher development capability of the algorithm is required. The results in Tables 2 and

**Table 1** The PR values of the 14 algorithms in CEC2013 when  $\epsilon = 1.0e-01$

Func	R2PSO	R3PSO	R2PSOLHC	R3PSOLHC	CDE	NCDE	SDE	FERPSO	NSDE	NSHDE	MO_Ring_PSO_SCD	ADE	SaEFSDE	SGD-DE
F1	0.475	0.988	0.9	0.925	<b>1</b>	<b>1</b>	<b>1</b>	0.941	textbf1	<b>1</b>	<b>1</b>	0.975	0.522	<b>1</b>
F2	0.2	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.84	<b>1</b>	<b>1</b>	<b>1</b>	0.731	<b>1</b>	<b>1</b>	<b>1</b>
F3	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
F4	0.25	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.794	0.258	<b>1</b>	<b>1</b>
F5	0.5	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.988	<b>1</b>	0.925	0.506	<b>1</b>	<b>1</b>
F6	0.024	0.976	0.967	0.979	0.999	<b>1</b>	0.956	0.999	<b>1</b>	<b>1</b>	0.392	0.245	0.932	0.944
F7	0.023	0.922	0.994	0.965	0.942	0.965	0.534	0.545	0.605	<b>1</b>	0.109	0.176	0.293	0.397
F8	0	0.009	0.008	0.011	<b>1</b>	<b>1</b>	0.134	0.426	0.833	<b>1</b>	0.094	0.028	0.303	0.287
F9	0	0.594	0.638	0.603	0.551	0.516	0.158	0.212	0.224	<b>0.795</b>	0.023	0.013	0.093	0.106
F10	0.083	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.563	0.992	0.992	<b>1</b>	0.347	0.122	0.79	0.988
F11	0.142	0.983	<b>1</b>	<b>1</b>	0.921	0.796	0.671	0.724	0.917	0.711	0.427	0.167	0.65	<b>1</b>
F12	0	0.619	0.647	0.584	0.338	0.784	0.934	0.836	0.984	0.753	0.304	0.125	0.328	0.893
F13	0.021	0.775	0.792	<b>0.829</b>	0.763	0.688	0.667	0.667	0.683	0.683	0.38	0.167	0.596	0.786
F14	0.004	0.304	0.271	0.321	<b>0.813</b>	0.7	0.646	0.65	0.713	0.688	0.388	0.167	0.636	0.667
F15	0	0.091	0.106	0.088	0.828	0.488	0.456	<b>0.888</b>	0.559	0.459	0.194	0.125	0.149	0.696
F16	0	0	0	0	0.646	<b>0.875</b>	0.429	0.742	0.783	0.81	0.288	0.167	0.299	0.667
F17	0	0	0	0	0.263	0.4	0.153	<b>1</b>	0.309	0.367	0.288	0.125	0.145	0.571
F18	0	0	0	0	0.75	0.9	0.15	<b>0.975</b>	0.576	0.908	0.163	0.167	0.186	0.69
F19	0	0	0	0	0	0.378	0.11	0	0.31	0.336	0.013	0.093	0.125	<b>0.482</b>
F20	0	0	0	0	0	0.419	0	0	<b>0.856</b>	0.339	0	0.125	0.125	0.25
Best	0	5	6	7	8	9	4	7	6	<b>10</b>	2	2	4	7



**Table 2** PR values of 14 algorithms in CEC2013 when  $\varepsilon = 1.0e-02$

	R2PSO	R3PSO	R2PSOLHC	R3PSOLHC	CDE	NCDE	SDE	FERPSO	NSDE	NSHDE	MO_Ring_PSO_SCD	ADE	SaEFSDE	SGD-DE
F1	0	0.35	0.163	0.375	1	1	1	0.125	1	1	0.394	0.975	0.522	1
F2	0.2	1	1	1	1	1	0.84	1	1	1	0.731	1	1	1
F3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F4	0.25	1	1	0.925	1	1	1	1	1	1	0.794	0.258	0.963	1
F5	0.5	1	1	0.888	0.913	0.888	1	0.925	0.925	0.969	0.925	0.506	1	1
F6	0.007	0.907	0.915	0.932	0.864	1	0.956	0.997	1	0.985	0.392	0.218	0.83	0.944
F7	0	0.576	0.602	0.603	0.912	0.917	0.534	0.544	0.604	0.965	0.109	0.175	0.293	0.389
F8	0	0.002	0.002	0.004	0.998	1	0.133	0.367	0.831	0.998	0.094	0.004	0.129	0.287
F9	0	0.222	0.228	0.227	0.549	0.515	0.15	0.211	0.224	0.647	0.023	0.013	0.093	0.102
F10	0.075	1	0.996	1	1	1	0.563	0.992	0.992	1	0.347	0.122	0.79	0.988
F11	0	0.667	0.671	0.667	0.488	0.7	0.671	0.675	0.849	0.684	0.427	0.167	0.555	0.976
F12	0	0.456	0.509	0.531	0.072	0.538	0.934	0.759	0.972	0.563	0.304	0.125	0.288	0.893
F13	0	0.667	0.658	0.667	0.283	0.667	0.667	0.667	0.667	0.667	0.38	0.167	0.522	0.786
F14	0	0.058	0.104	0.146	0.263	0.667	0.646	0.65	0.667	0.667	0.388	0.167	0.515	0.667
F15	0	0.031	0.025	0.022	0.109	0.369	0.391	0.422	0.469	0.33	0.194	0.125	0.141	0.696
F16	0	0	0	0	0.05	0.667	0.429	0.617	0.658	0.665	0.288	0.167	0.192	0.667
F17	0	0	0	0	0	0.263	0.141	0.197	0.25	0.292	0.128	0.125	0.125	0.536
F18	0	0	0	0	0.108	0.35	0.15	0.157	0.396	0.376	0.146	0.167	0.167	0.643
F19	0	0	0	0	0	0.238	0.013	0	0.248	0.26	0	0.093	0.125	0.482
F20	0	0	0	0	0	0.125	0	0	0.031	0.133	0	0.125	0.125	0.25
Best	0	5	4	5	4	9	3	4	6	8	1	1	2	14

**Table 3** PR values of 14 algorithms in CEC2013 when  $\varepsilon = 1.0e-05$

	R2PSO	R3PSO	R2PSOLHC	R3PSOLHC	CDE	NCDE	SDE	FERPSO	NSDE	NSHDE	MO_Ring_PSO_SCD	ADE	SaEFSDE	SGD-DE
F1	0	0.35	0.163	0.375	1	1	1	0.125	1	1	0.394	0.975	0.522	1
F2	0.2	1	1	1	1	1	0.84	1	1	1	0.731	1	1	1
F3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F4	0.25	1	1	1	0.925	1	1	1	1	1	0.794	0.258	0.963	1
F5	0.5	1	1	1	0.888	0.913	0.888	1	0.925	0.969	0.925	0.506	1	1
F6	0.007	0.907	0.915	0.932	0.864	1	0.956	0.997	1	0.985	0.392	0.218	0.83	0.944
F7	0	0.576	0.602	0.603	0.912	0.917	0.534	0.544	0.604	0.965	0.109	0.175	0.293	0.389
F8	0	0.002	0.002	0.004	0.998	1	0.133	0.367	0.831	0.998	0.094	0.004	0.129	0.287
F9	0	0.222	0.228	0.227	0.549	0.515	0.15	0.211	0.224	0.647	0.023	0.013	0.093	0.102
F10	0.075	1	0.996	1	1	1	0.563	0.992	0.992	1	0.347	0.122	0.79	0.988
F11	0	0.667	0.671	0.667	0.488	0.7	0.671	0.675	0.849	0.684	0.427	0.167	0.555	0.976
F12	0	0.456	0.509	0.531	0.072	0.538	0.934	0.759	0.972	0.563	0.304	0.125	0.288	0.893
F13	0	0.667	0.658	0.667	0.283	0.667	0.667	0.667	0.667	0.667	0.38	0.167	0.522	0.786
F14	0	0.058	0.104	0.146	0.263	0.667	0.646	0.65	0.667	0.667	0.388	0.167	0.515	0.667
F15	0	0.031	0.025	0.022	0.109	0.369	0.391	0.422	0.469	0.33	0.194	0.125	0.141	0.696
F16	0	0	0	0	0.05	0.667	0.429	0.617	0.658	0.665	0.288	0.167	0.192	0.667
F17	0	0	0	0	0	0.263	0.141	0.197	0.25	0.292	0.128	0.125	0.125	0.536
F18	0	0	0	0	0.108	0.35	0.15	0.157	0.396	0.376	0.146	0.167	0.167	0.643
F19	0	0	0	0	0	0.238	0.013	0	0.248	0.26	0	0.093	0.125	0.482
F20	0	0	0	0	0	0.125	0	0	0.031	0.133	0	0.125	0.125	0.25
Best	0	5	4	5	4	9	3	4	6	8	1	1	2	14

**Table 4** Influence of local search on algorithm accuracy

Func.	Local search	Without local search
Average	– <b>0.027294278</b>	– 0.027973988
Standar	<b>5.57037E–05</b>	0.001090751

3 show that the gradient descent based local search strategy proposed in this paper effectively improves the exploitation ability of the algorithm, and the optimal solution found at lower precision can improve the algorithm’s search accuracy (Table 4), which is shown in Table 4 where bolded font indicate better search accuracy.

**Effect of the mutation strategy on performance**

The niche technology are used in SGD-DE to generate populations, and the local search strategy of gradient descent is utilized to further exploit the solution space between the niches while sub-populations communicate, thus enhancing the searching ability of the algorithm. Without the local search strategy, the algorithm is just a simple species-based DE. Table 1 shows that the performance of the proposed algorithm is significantly superior to that of Species-based DE; Without the niche technology, the algorithm degenerates into a traditional DE algorithm with better exploration capabilities, but it can still get trapped in the local optima. Therefore, those two improved strategies are indispensable for SGD-DE.

**Effect of the selection strategy on performance**

In this section, CEC2013 dataset is used to test the effectiveness of our hybrid selection strategy. The two algorithms used for comparison are the SGD-DE with elite strategy, and the SGD-DE with crowding strategy. For convenience, we only put results of the lowest accuracy and the highest accuracy.

Tables 5 and 6 describe the experimental results of SGD-DE using the hybrid selection strategy and the other two single selection strategies, with the bolded font indicating the better experimental results. It can be seen that SGD-DE is better than the other two algorithms overall, especially better than the SGD-DE with single elite strategy, which is because the crowding strategy can better ensure the population diversity so that shows better performance on F6, F7, and F8 when solving the optimal problems, but the results of SGD-DE with a mixed selection strategy on these problems are almost the same as SGD-DE with single crowding strategy. Meanwhile, SGD-DE with single crowding strategy also performs well on the low-dimensional functions.

**Table 5** Experimental results in PR on CEC 2013 problems at accuracy level  $\varepsilon = 1.0e-01$

Func.	SGD-DE	Elite	Crowd
F1	<b>1</b>	0.357	<b>1</b>
F2	<b>1</b>	<b>1</b>	<b>1</b>
F3	<b>1</b>	<b>1</b>	<b>1</b>
F4	<b>1</b>	0.643	<b>1</b>
F5	<b>1</b>	<b>1</b>	<b>1</b>
F6	0.944	0.397	<b>0.984</b>
F7	0.397	0.5	<b>0.444</b>
F8	0.287	0.044	<b>0.37</b>
F9	<b>0.106</b>	0.122	0.098
F10	0.988	0.595	<b>1</b>
F11	<b>1</b>	0.024	0.952
F12	<b>0.893</b>	0.054	0.911
F13	<b>0.786</b>	0.071	0.738
F14	0.667	0.048	<b>0.69</b>
F15	<b>0.696</b>	0	0.607
F16	<b>0.667</b>	0.024	<b>0.667</b>
F17	<b>0.571</b>	0.054	0.536
F18	<b>0.69</b>	0.286	<b>0.69</b>
F19	<b>0.482</b>	0	<b>0.482</b>
F20	0.25	<b>0.268</b>	0
BEST	1	1	1

**Table 6** Experimental results in PR on CEC 2013 problems at accuracy level  $\varepsilon = 1.0e-05$

Func.	SGD-DE	Elite	Crowd
F1	<b>1</b>	0.357	<b>1</b>
F2	<b>1</b>	0.314	<b>1</b>
F3	<b>1</b>	0.571	<b>1</b>
F4	<b>1</b>	0.036	<b>1</b>
F5	<b>1</b>	0.143	<b>1</b>
F6	0.944	0.095	<b>0.984</b>
F7	0.397	0.202	<b>0.444</b>
F8	0.287	0.011	<b>0.37</b>
F9	<b>0.106</b>	0.056	0.098
F10	<b>0.988</b>	0.012	<b>0.988</b>
F11	<b>1</b>	0	0.905
F12	<b>0.893</b>	0	0.893
F13	<b>0.786</b>	0	0.714
F14	<b>0.667</b>	0	<b>0.667</b>
F15	<b>0.696</b>	0	0.554
F16	<b>0.667</b>	0	<b>0.667</b>
F17	<b>0.571</b>	0	0.5
F18	<b>0.69</b>	0.024	0.476
F19	<b>0.482</b>	0	0.393
F20	<b>0.25</b>	0.214	0
BEST	1	0.571	1

**Table 7** Experimental results of PR on niche size parameter at the accuracy level  $\varepsilon = 1.0e-04$

Func.	NS = 20	NS = 50
F1	1	1
F2	1	1
F3	0.429	1
F4	0.929	1
F5	1	1
F6	0.532	0.944
F7	0.278	0.377
F8	0.108	0.286
F9	0.042	0.092
F10	0.69	0.988
F11	0.738	0.976
F12	0.518	0.893
F13	0.619	0.762
F14	0.667	0.667
F15	0.589	0.696
F16	0.643	0.667
F17	0.393	0.5
F18	0.524	0.643
F19	0.304	0.464
F20	0.25	0.25

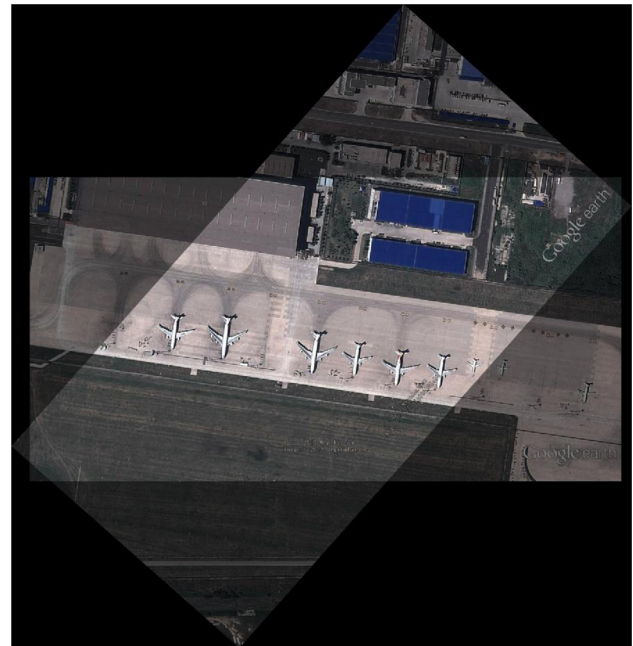
### Effect of the niche size on performance

SGD-DE is mainly affected by the niche size. In this section, we tested the effects of different niche sizes on CEC2013 test results. The results of NS = 20 and NS = 50 in CEC2013 test set are tested. Table 7 shows that different niche sizes have great influence on the results in different functions.

### Application for image registration

#### Registration results

Firstly, the proposed method is applied to the registration problem with images that have little difference. Dataset of Object Detection in Aerial Images from the University of



**Fig. 5** The registration results for the UCAS-AOD dataset

Chinese Academy of Sciences (UCAS-AOD) are tested, and the deviation of the images is mainly translation and rotation. The UCAS-AOD data set is shown in Fig. 4, and the image registration result of UCAS-AOD registration is shown in Fig. 5.

the above two images overlapped after registration, and the corresponding optimal solutions can be found with our proposed algorithm.

Then the aerial images from Guangxi University are tested to detect the registration performance of our method on images that have large differences. Figures 6 and 7 show the images shot at different times and devices, the registration result is shown in Fig. 8.

The registration results of two images of Guangxi University shot in different years show that the aligned image produces almost no ghosting, and the corresponding optimal solutions can still be found with the algorithm proposed



**Fig. 4** The UCAS-AOD data set



**Fig. 6** The image of Guangxi University in 2019



**Fig. 7** The image of Guangxi University in 2018



in this paper even if there are some significant differences between two images.

### Registration convergence speed

In the matching of remote sensing images of Guangxi University, this algorithm is compared with the traditional optimization algorithm and the current advanced multi-modal optimization algorithm such as DE, and four PSO algorithms using ring topology [14].

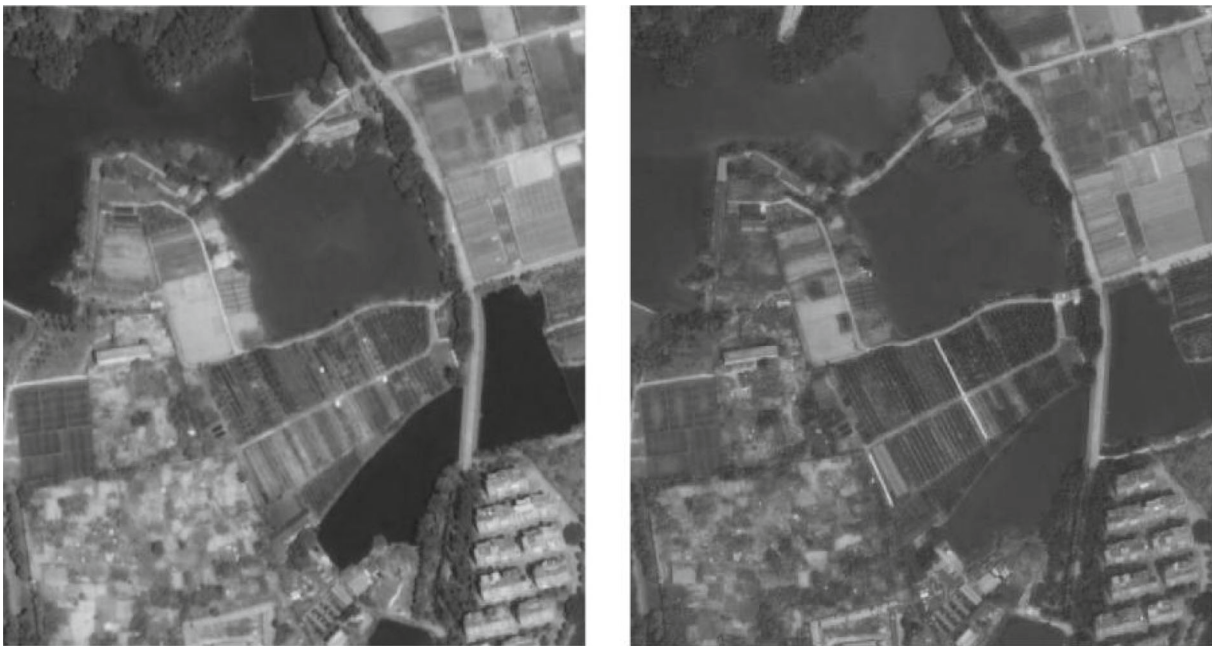
Figure 9 shows that the DE algorithm is easy to fall into the local optimum when registering complex image problems. SGD-DE converges earliest and fastest at the beginning, and when trapped in a local optimum, it takes less iterations to jump out of the local optimum, and SGD-DE is more exploitable and exploratory. Our algorithm is faster and more accurate in image registration than the advanced multi-modal PSO algorithm. Because we not only use niche technique

to increase diversity but also add local search strategy to improve accuracy.

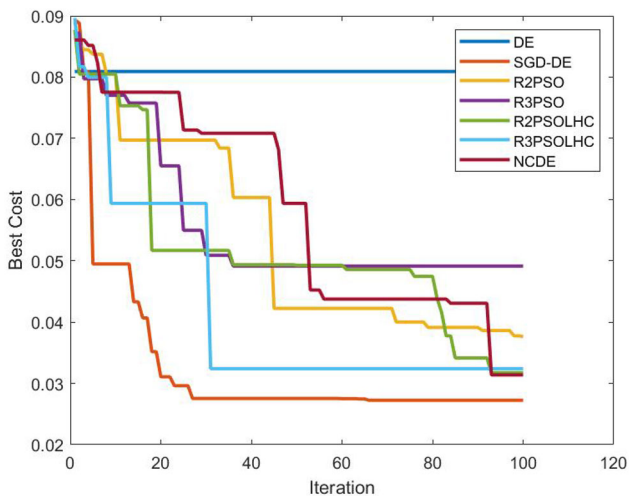
### Registration accuracy

Compare the DTV function value of SGD-DE with R2PSO, R3PSO, R2PSOLHC, R3PSOLHC, NCDE, FERPSO and NSDE after each image registration experiment. According to the definition of the function, the smaller the DTV value, the smaller the difference between the overlapping parts of the image, and the more accurate the registration result will be.

In this experiment, it mainly compares the search capability of the algorithm and the ability to avoid falling into local optimum. From the image of the DTV function, it can be seen that the function has many local optima, and the optimization algorithm may stagnate. As is shown in Table 8, the bolded font indicate better mean value and standard deviation. It can be seen from the average and standard deviation of the exper-



**Fig. 8** The result of remote sensing image registration of Guangxi University



**Fig. 9** Iterative data of various algorithms

iment results that SGD-DE is superior to other multi-modal optimization algorithms.

In experiment one and experiment two, we tested the SGD-DE algorithm that optimizes the NMI function and the SGD-DE algorithm that optimizes the DTV function. And the NMI value and DTV value of the two registration results are compared. From the comparison results we can see which image registration function is more accurate.

DTV is the minimization function and NMI is the maximum function. It can be seen from the table that the registration result optimized by DTV function is more accurate than other methods. Therefore, it is concluded that the

DTV method is more precise and robust, and the image registration results are described in Table 9.

Among them, DTV is the minimization function, and NMI is the maximization function. It can be seen from the table that the accuracy of the registration result optimized by the DTV function is higher than that of other methods, so the DTV method is more accurate and robust.

## Conclusion

This paper proposed an SGD-DE for tackling the remote sensing image registration problem, which enhances the capability of exploration and exploitation. Experiments have shown that our algorithm can achieve a promising performance when finding the optimal solutions to the remote sensing image registration problem, and the registration result is relatively accurate. In terms of some performance indicators, our algorithm is superior to the current advanced ones. At present, the algorithm still needs improvement, such as niche division and parameter adaptation, so we will further improve our algorithm with the clustering method in the future.

In the future research, we would like to do further research on metrics for comparison of rendered and reference images, focusing on topological similarities between the phenotype and a reference image (e.g. number of subbranches and their lengths). Multiple metrics could be used and combined with the use of multi-objective search, and possibly combined with interactive methods for optimization. One of the ideas for

**Table 8** Mean value and standard deviation of DTV function

	R2PSO	R3PSO	R2PSOLHC	R3PSOLHC	NCDE	FERPSO	NSDE	SGD-DE
MEAN	− 0.03174	− 0.03635	− 0.05502	− 0.04362	− 0.03144	− 0.05764	− 0.02741	− <b>0.02729</b>
STD	0.0037	0.007115	0.012543	0.014163	0.006143	0.021013	6.70E−05	<b>5.57E−05</b>

**Table 9** Image registration results of Guangxi University

	DTV		NMI	
	Mean	Best	Mean	Best
SGD-DE with DTV	0.027337905	0.027246878	1.18042428	1.178915963
SGD-DE with NMI	0.029343468	0.028543298	1.17745289	1.169815963

future research may also include another encoding aspect of the procedural model using evolution of line segments for vector parameters. In our future work, we aim to apply the proposed algorithm to improve the capability of finding optimal feasible solutions in large scale problems.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (Grant no. 61763002), Foundation of Guangxi Experiment Center of Information Science (Grant no. KF1401) and Natural Science Foundation of Guangxi Zhuang Autonomous Region (2018GXNSFAA294133).

## Declaration

**Conflict of interest** We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled. On behalf of all authors, the corresponding author states that there is no conflict of interest.

**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/>.

## References

- Araújo RL, Ushizima DM, Silva RR (2020) Fusion of color bands using genetic algorithm to segment melanoma. In: 2020 IEEE 17th international symposium on biomedical imaging workshops (ISBI Workshops). IEEE, pp 1–4
- Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (surf). *Comput Vis Image Underst* 110(3):346–359
- Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Mach Intell* 24(4):509–522
- Cheung K, Siu Y, Shen T (2019) Fast adaptive bases algorithm for non-rigid image registration. *J Imaging Sci Technol* 63(1):10505–10511
- Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Trans Evol Comput* 6(2):182–197. <https://doi.org/10.1109/4235.996017>
- Fan Q, Yan X (2015) Self-adaptive differential evolution algorithm with zoning evolution of control parameters and adaptive mutation strategies. *IEEE Trans Cybern* 46(1):219–232
- Gong M, Zhao S, Jiao L, Tian D, Wang S (2013) A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information. *IEEE Trans Geosci Remote Sens* 52(7):4328–4338
- Gottesfeld Brown L (1992) A survey of image registration techniques. *Acm Comput Surv* 24(4):325–376
- Hisham M, Yaakob SN, Raof RA, Nazren AA, Embedded NW (2015) Template matching using sum of squared difference and normalized cross correlation. In: 2015 IEEE student conference on research and development (SCoReD). IEEE, pp 100–104
- Iacca G, Caraffini F, Neri F (2015) Continuous parameter pools in ensemble differential evolution. In: 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, pp 1529–1536. <https://doi.org/10.1109/SSCI.2015.216>
- Li R, Peng Y, Shi H, Wu H, Liu S, Kwok N (2019) First-order difference bare bones particle swarm optimizer. *IEEE Access* 7:132472–132491
- Li X (2005) Efficient differential evolution using speciation for multimodal function optimization. In: Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation, Association for Computing Machinery, New York, USA, pp 873–880. <https://doi.org/10.1145/1068009.1068156>
- Li X (2007) A multimodal particle swarm optimizer based on fitness euclidean-distance ratio. In: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, Association for Computing Machinery, New York, USA, pp 78–85. <https://doi.org/10.1145/1276958.1276970>
- Li X (2009) Niching without niching parameters: particle swarm optimization using a ring topology. *IEEE Trans Evol Comput* 14(1):150–169
- Li X, Engelbrecht A, Eptropakis MG (2013) Benchmark functions for cec'2013 special session and competition on niching methods for multimodal function optimization. RMIT University, Evolutionary Computation and Machine Learning Group, Australia, Tech Rep



16. Li Y, Chen C, Yang F, Huang J (2015) Deep sparse representation for robust image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, pp 4894–4901. <https://doi.org/10.1109/CVPR.2015.7299123>
17. Li Y, Chen C, Zhou J, Huang J (2015) Robust image registration in the gradient domain. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, pp 605–608
18. Liu L, Piao C, Jiang X, Zheng L (2018) Research on governmental data sharing based on local differential privacy approach. In: 2018 IEEE 15th international conference on e-Business engineering (ICEBE). IEEE, pp 39–45
19. Liu S, Yan X, Li P, Hao X, Wang K (2018) Radar emitter recognition based on sift position and scale features. *IEEE Trans Circuits Syst II Express Briefs* 65(12):2062–2066
20. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
21. Matindife L, Sun Y, Wang Z (2021) Image-based mains signal disaggregation and load recognition. *Complex Intell Syst* 7(2):901–927
22. Onofrey JA, Papademetris X, Staib LH (2015) Low-dimensional non-rigid image registration using statistical deformation models from semi-supervised training data. *IEEE Trans Med Imaging* 34(7):1522–1532. <https://doi.org/10.1109/TMI.2015.2404572>
23. Poláková R, Tvrdík J, Bujok P (2019) Differential evolution with adaptive mechanism of population size according to current population diversity. *Swarm Evol Comput* 50:100519
24. Preuss M (2015) Niching methods and multimodal optimization performance. In: Multimodal optimization by means of evolutionary algorithms. Springer International Publishing, Cham, pp 115–137
25. Qu BY, Suganthan PN, Liang JJ (2012) Differential evolution with neighborhood mutation for multimodal optimization. *IEEE Trans Evol Comput* 16(5):601–614
26. Raz AK, Llinas J, Mittu R, Lawless WF (2020) Engineering for emergence in information fusion systems: a review of some challenges. *Hum Mach Shared Contexts* 241–255
27. Rybintsev A (2017) Age estimation from a face image in a selected gender-race group based on ranked local binary patterns. *Complex Intell Syst* 3(2):93–104
28. Schmitt M, Zhu XX (2016) Data fusion and remote sensing: an ever-growing relationship. *IEEE Geosci Remote Sens Mag* 4(4):6–23
29. Shen D, Luo S (2018) Crowding-based differential evolution with self-adaptive control parameters for dynamic environments. In: 2018 14th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, pp 71–76
30. Storn R, Price K (1997) Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359
31. Sun J, Chen X, Zhang J, Yao W (2021) A niching cross-entropy method for multimodal satellite layout optimization design. *Complex Intell Syst* 1–19
32. Thomsen R (2004) Multimodal optimization using crowding-based differential evolution. In: Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), vol 2. IEEE, pp 1382–1389
33. Ursem RK (1999) Multinational evolutionary algorithms. In: Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol 3. IEEE, pp 1633–1640
34. Wang F, Zhang H, Li K, Lin Z, Yang J, Shen XL (2018) A hybrid particle swarm optimization algorithm using adaptive learning strategy. *Inf Sci* 436:162–177
35. Wang F, Li Y, Zhou A, Tang K (2019) An estimation of distribution algorithm for mixed-variable newsvendor problems. *IEEE Trans Evol Comput* 24(3):479–493
36. Wang F, Li Y, Liao F, Yan H (2020) An ensemble learning based prediction strategy for dynamic multi-objective optimization. *Appl Soft Comput* 96:106592
37. Wang F, Zhang H, Zhou A (2020) A particle swarm optimization algorithm for mixed-variable optimization problems. *Swarm Evol Comput* 60:100808
38. Wang H, Wang W, Zhou X, Zhao J, Wang Y, Xiao S, Xu M (2020) Artificial bee colony algorithm based on knowledge fusion. *Complex Intell Syst* 1–14
39. Wang ZJ, Zhan ZH, Zhang J (2019) Distributed minimum spanning tree differential evolution for multimodal optimization problems. *Soft Comput* 23(24):13339–13349
40. Yang Y, Duan Z (2020) An effective co-evolutionary algorithm based on artificial bee colony and differential evolution for time series predicting optimization. *Complex Intell Syst* 6:299–308
41. Yue C, Qu B, Liang J (2017) A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems. *IEEE Trans Evol Comput* 22(5):805–817
42. Zhang W, Li G, Zhang W, Liang J, Yen GG (2019) A cluster based PSO with leader updating mechanism and ring-topology for multimodal multi-objective optimization. *Swarm Evol Comput* 50:100569
43. Zhang Y, Zhang M (2020) Machine learning model-based two-dimensional matrix computation model for human motion and dance recovery. *Complex Intell Syst* 1–11
44. Zhao F, Huang Q, Gao W (2006) Image matching by normalized cross-correlation. In: 2006 IEEE international conference on acoustics speech and signal processing proceedings, vol 2. IEEE, pp II
45. Zhao H, Zhan ZH, Lin Y, Chen X, Luo XN, Zhang J, Kwong 800 S, Zhang J (2019) Local binary pattern-based adaptive differential evolution for multimodal optimization problems. *IEEE Trans Cybern* 50(7):3343–3357

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