Swarm and Evolutionary Computation 42 (2018) 138–159

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

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

Coral Reef Optimization with substrate layers for medical Image Registration

Enrique Bermejo [a], Manuel Chica [b], Sergio Damas [c], Sancho Salcedo-Sanz [d], Oscar Cordón [a,e,*]

[a] Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, Spain
[b] School of Electrical Engineering and Computing, The University of Newcastle, 2308 Callaghan, NSW, Australia
[c] Department of Software Engineering, University of Granada, Granada 18071, Spain
[d] Department of Signal Theory And Communications, University of Alcalá, Alcalá de Henares 28805, Spain
[e] Research Center on Information and Communication Technologies, University of Granada, Granada 18071, Spain

ARTICLE INFO

Keywords:
Image Registration
Medical imaging
Coral Reef Optimization
Nature-inspired algorithm
Swarm intelligence

ABSTRACT

In medical imaging there is a special interest in relating information from different images frequently used for diagnosis or treatment. Image registration (IR) involves the transformation of different sets of image data having a shared content into a common coordinate system. The estimation of the optimal transformation is modelled either as a combinatorial or a numerical optimization problem. Since traditional IR methods are constrained by several limitations, other optimization methods have been recently proposed to overcome such shortcomings. In this contribution, we consider the use of a recently proposed and high performance bio-inspired meta-heuristic: the Coral Reef Optimization Algorithm with Substrate Layers (CRO-SL). We adapt the algorithm to the real-coding IR problem variant following both feature-based and intensity-based designs, and perform two thorough experimental studies. Such studies focus on both mono-modal and inter-modal scenarios where the images suffer different types of 3D affine transformations to validate our proposal. The new proposal is benchmarked with state-of-the-art evolutionary and non-evolutionary IR methods. The results show that CRO-SL is a very competitive approach in terms of its robustness, accuracy, and efficiency.

1. Introduction

Image analysis, understanding, and visualization are essential tasks in medical and scientific applications. Many of these applications require the comparison, integration, or fusion of visual information acquired from different imaging devices, at different times, or distinct viewpoints. In particular, medical images are used for diagnosis and disease monitoring, intervention and treatment planning, assisted surgery, comparison of the patient's data with anatomical atlases, and anatomy segmentation, among many other applications. Generally, these images are acquired using multiple devices (i.e., Computer Tomography (CT), X-ray, Magnetic Resonance Imaging (MRI), Positron Emission Tomography, Ultrasound, etc.), and involve different acquisition moments (before/after treatment). Misalignment between images is thus inevitable, regardless they belong to a single or multiple patients (cohort studies). Therefore, Image Registration (IR) [1] is an essential preprocessing task in medical imaging, as it allows the alignment and integration of visual information obtained under different conditions.

Specifically, IR aims to estimate the optimal *geometric transformation* that overlaps the common part of the images. In most IR approaches, the objective is to transform an input image to reach the geometry of a reference image. Both images are related by a spatial transformation and a *similarity metric* is used to measure the degree of resemblance between them [2]. The problem of finding the optimal transformation is tackled by an iterative optimization procedure that explores the space of possible transformations guided by the similarity metric.

IR methods can be divided into two major categories: feature-based, where salient or distinctive parts of the images are used to register the images, and intensity-based, if the alignment procedure makes use of the entire image. Feature-based approaches [1] aim to expedite the optimization process, reducing the complexity of the problem by only using a portion of the images. These methods are notably reliant on

* Corresponding author. Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, Spain.
E-mail addresses: enrique.bermejo@decsai.ugr.es (E. Bermejo), manuel.chicaserrano@newcastle.edu.au (M. Chica), sdamas@ugr.es (S. Damas), sancho.salcedo@uah.es (S. Salcedo-Sanz), ocordon@decsai.ugr.es (O. Cordón).

https://doi.org/10.1016/j.swevo.2018.03.003
Received 7 March 2017; Received in revised form 21 December 2017; Accepted 2 March 2018
Available online 6 March 2018
2210-6502/

the technique used to extract the feature points, which is usually robust to noise and distortion. However, this approach is only suitable if features are able to provide representative information of the images. On the other hand, intensity-based (also known as voxel-based) approaches can deal with a larger amount of data at the expense of increasing the computational requirements. The alignment is guided by the distribution of intensity values (gray levels) of the whole images, which makes these approaches more precise but sensitive to intensity changes when illumination changes and noise are present.

A wide variety of applications have been proposed in the medical IR field. The iterative Closest Point algorithm [3], the original Powell's method [4] or some more recent variants [5], different gradient-based approaches [6], and discrete optimization of Markov random field (MRF) formulation [7] are well-known classical optimization techniques. Nevertheless, aspects such as presence of noise, image discretization and misalignment, among others, negatively affect the optimization process of traditional methods, which are prone to be trapped in local minima. IR methods based on evolutionary algorithms, hybridizations, and other metaheuristics (MHs) [8] have demonstrated to overcome some of these drawbacks. Although they are not free of the local minima issue, MHs are usually considered as global optimization techniques due to their robust performance, which makes them suitable to face complex medical IR problems. Several evolutionary methods have been so far proposed to deal with different IR problems improving the results obtained by traditional methods [9,10]. Among them, an approach based on scatter search (SS) [11] provided outstanding and consistent results in many comparisons [12,13].

Recently, a novel bio-inspired evolutionary-type MH called Coral Reef Optimization (CRO) algorithm was proposed in Ref. [14], and later improved in Ref. [15] with the integration of a competitive co-evolution scheme, known as CRO with substrate layers (CRO-SL). CRO and CRO-SL are evolutionary algorithms based on the artificial simulation of the natural phenomena taking place in the formation and reproduction of coral reefs. During their life, corals undergo different phases, such as reproduction, larval settlement, or fight for a space in the reef where they can survive. The proposed CRO-based approaches emulate these processes favoring a powerful trade-off between diversity and specificity, which makes them suitable for tackling complex optimization problems.

Since CRO-SL has been successfully applied to different real-world problems demonstrating a robust performance, we consider it can also be applied to complex IR problems and behave adequately. Thus, in the present work, we design novel IR methods based on CRO and CRO-SL where both the objective function and the coding scheme are adapted to a specific medical IR problem.

In order to validate our proposal, we develop an exhaustive experimental setup comparing both CRO and CRO-SL against some state-of-the-art IR methods in different feature- and intensity-based medical IR scenarios. Two sets of problems were considered. First, a preliminary mono-modal study where each scenario has been generated using MRI scans extracted from the well-known BrainWeb database at McGill University [16]. Second, a multi-modal study (i.e. images acquired using different sensors) where each scenario consists of pairs of CT and MRI scans extracted from the Retrospective Image Registration Evaluation Project [17]. Different statistical tests are applied to evaluate the significance of the results.

There are two main reasons that justify such two broad experimental studies. First, we aim to compare simultaneously evolutionary feature-based and intensity-based IR approaches and we thus consider the same conditions of previous contributions [12,13], i.e. a mono-modal study. Second, there are rather different CRO-based designs and that preliminary mono-modal study aims to determine the best performing CRO-based configuration. Once tested and analyzed in a mono-modal environment, the best performing and robust evolutionary methods will later face a more challenging multi-modal environment including more complex transformations and other non-evolutionary methods.

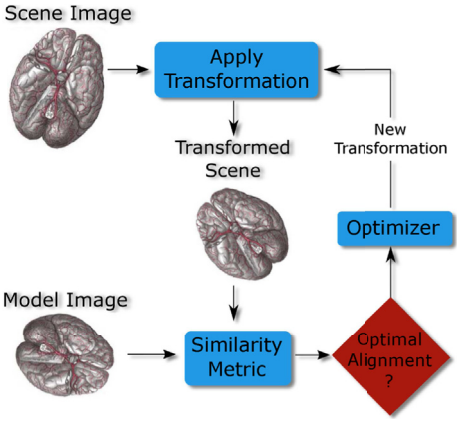

Fig. 1. The interactions among the three components (blue boxes) of a registration technique. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The paper is structured as follows. Sections 2 and 3 introduce the basics of IR, detail and review the most relevant IR methods, focusing specifically on the 3D medical IR problem. Section 4 describes the basics of the CRO algorithm and the peculiarities of the enhanced CRO-SL variant regarding different exploratory operators. Their adaptation to solve an IR problem is presented in Section 5. Section 6 introduces the multi-modal test scenarios, the experimental design considering outstanding intensity-based approaches, and a detailed analysis of the results. Section 7 draws some conclusions and remarks on the research carried out. Additionally, two appendices have been included to complement the analysis and provide a comprehensive view of the behavior of our proposal in terms of optimization performance. Appendix A introduces the mono-modal test scenarios, the specific experimental designs for feature- and intensity-based evolutionary approaches, a preliminary analysis of the substrate configuration for CRO-SL, and an exhaustive analysis of the results obtained in the experimentation developed. Last, Appendix B summarizes the results of a simplified version of the multi-modal study considering rigid body transformations.

2. Image registration

2.1. Problem statement

In general, the registration framework is defined as an optimization problem that aims to maximize the alignment between two input images: a reference image, known as *model*, and a *scene* image, which is transformed to reach the model geometry [1]. In other words, the goal is to find a geometric transformation that is applied over the scene in order to maximize its similarity with the model. There are many considerations that must be taken into account when designing any IR method, as each application has its own specific characteristics. In particular, there are three essential components involved in a typical IR method: the transformation model, the similarity metric, and the optimization procedure (see Fig. 1).

Fig. 2 presents an example of the application of different *transformation models* to a brain MRI scan. Most medical IR applications involving anatomical structures like bones or the brain [18] consider rigid transformations (allowing rotation and translation) to avoid unrealistic representations of the skeleton or tissue growth. If scaling is also required, a similarity transformation is then considered. Applications using CT scans may require non-uniform scaling due to the presence of shearing. In these cases, affine transformations are considered. When soft tissue is involved and deformation can be expected, deformable transformations are commonly used [19].

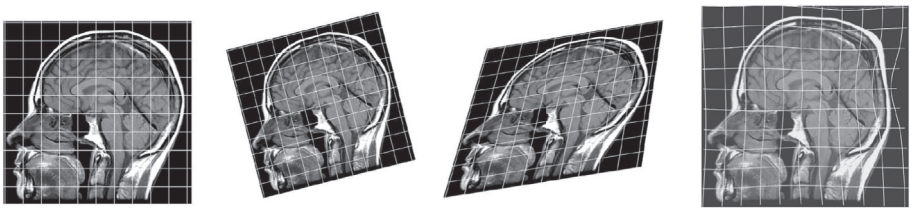

Fig. 2. From left to right, three images obtained from the same initial scene by applying different transformations: similarity, affine, and B-spline.

The *similarity metric* measures the quality of the alignment or the degree of resemblance between the transformed scene and the model. The choice of the similarity metric is a crucial step in the design of any IR method [20] and it depends on the considered registration approach. Feature-based methods usually consider the mean square error (MSE) as similarity metric to measure the distance between corresponding features [21]. As the intensity values of the two images is considered in intensity-based approaches, the relationship between intensity distributions is measured by the similarity metric to evaluate the alignment. There are multiple metrics available (e.g., sum of squared differences, normalized correlation (NC), mutual information (MI), among others) and their suitability is determined by the acquisition technique of the images and the inherent relationship between them.

The *optimizer* tries to refine the solution until a particular stopping criteria is reached (e.g., when a suitable solution has been found or the algorithm performs a determined number of iterations). Hence, the optimization of the spatial transformation involves an iterative process to explore the search space of the geometrical transformation. According to the characteristics of the search space, we can find two different strategies. When the search is performed in the space of the transformation parameters, the registration is handled as a continuous optimization problem by parameter-based approaches. Alternatively, the search can also be performed on the space of correspondences by matching features (feature-based) or areas of the image (intensity-based) to align the images.

3. Current outstanding medical IR approaches

In the literature there are numerous surveys comparing and analyzing different IR approaches, from traditional to novel evolutionary and MH-based IR methods [1,22]. Next, we detail the characteristics of different classical methods and recent evolutionary MHs in the medical IR state of the art [13,23].

3.1. Classical parameter-based IR approaches

Nelder–Mead method. Also known as Downhill Simplex [24], is a local nonlinear optimization method based on the concept of a simplex, that is a set of n+1 points in a n-d space used to compute the cost function. The simplex is then modified by applying a reflection, shrinkage, contraction or expansion during the optimization process until the function is minimized.

Klein et al.'s ASGD. Adaptive stochastic gradient descent (ASGD) [6] was specifically designed for intensity-based IR. This method is an extension of the Robbins–Monro stochastic gradient descent method where the step size parameter is automatically computed. The optimization procedure of ASGD integrates a random image sampler and a multi-resolution strategy in which both a down-sampling and a Gaussian smoothing are applied to create reduced representation of the original images. In addition, a restart mechanism has been included to exploit the stochastic properties of this method.

Jenkinson et al.'s FLIRT. FMRIB's Linear Image Registration Tool [25] is a global optimization method designed for affine IR. Essentially, FLIRT is based on the combination of Powell's local optimization method [4] and an initial search phase. The optimization procedure also integrates a multi-resolution strategy with four resolution levels. During the first level, FLIRT performs an initial grid search to estimate the rotation parameters. Then, different local optimizations are applied in the remaining levels introducing perturbations to avoid local minima and correct mis-registrations while estimating the rest of the parameters.

Rohlfing et al.'s CMTK. The registration tool included in the Computational Morphometry Toolkit (CMTK) [26] is composed of an intensity-based multi-resolution algorithm based on the method proposed by Studholme et al. [27]. Precisely, it considers a four-level multi-resolution strategy allowing different degrees of freedom (DOF) at each stage.

Zikic et al.'s MRF. In Ref. [7], the authors propose an intensity-based method based on discrete optimization. In particular, Markov Random Fields (MRFs) are used to discretize the parameter search space, reducing the computational requirements of the procedure by considering a combinatorial optimization problem and allowing a fast convergence of the algorithm. In addition, this method considers a multi-resolution strategy involving down-sampling and Gaussian smoothing.

3.2. Classical matching-based IR approaches

Besl et al.'s ICP. ICP [3] is an iterative gradient-based method based on the least squares estimation of the transformation parameters from the computed correspondence between the feature points of the images considering the closest assignment rule. Hence, the optimization is guided by point matching instead of a similarity metric.

3.3. Intensity-based IR approaches

Valsecchi et al.'s r-*GA**. r-GA* is based on a *genetic algorithm* [28]. The optimizer component follows a real coded design in which a solution stores the transformation parameters in a real vector. The genetic operators are also real coded, i.e., blend crossover (BLX-α) [29] and random mutation. r-GA* is able to handle different similarity metrics and transformation models. In addition, it integrates the use of multiple resolutions combined with a restart and a search space adaptation mechanism. These mechanisms aim to speed up the optimization process and mitigate a premature convergence of the algorithm.

Valsecchi et al.'s SS*. The optimization component of SS* [13] is a variant of the original *scatter search* design [30]. In this variant, the reference set is divided in two tiers, containing the most high quality and diverse solutions, respectively. The specific methods integrated in the SS template were specifically designed for IR [31]. Thus, SS* integrates a diversification generation method based on a frequency memory [11], the solution combination method is the BLX-α crossover, and the improvement method is based on the PMX-α [32] operator. The reference set update method is responsible for classifying the solutions according to their quality and their diversity, and for maintaining the best ones of each category in the reference set. Last, the duplication control method prevents the appearance of identical copies of a solution in the reference set.

3.4. Feature-based evolutionary IR approaches

He and Narayana's HE-GA. This method was proposed in Ref. [9] and considers a real coding genetic scheme combined with dividing rectangle, a deterministic global optimization method based on branch

and bound. The GA follows an elitist generational model making use of arithmetic crossover and uniform mutation operators. It is applied in a first stage, where the GA estimates a preliminary solution, exploring a large search space. Then, the dividing rectangle method is applied to refine this solution by means of a local search procedure. A restart mechanism is also used in case of premature convergence.

Santamaría et al.'s SS. This SS-based IR method was proposed in Ref. [33] and it was originally designed to tackle range IR problems. The authors adopted a similarity metric based on the median square error (MedSE), specifically designed for images with low overlapping due to its robustness. They also integrated a grid closest point data structure [34] in order to speed up the computation of the fitness function, and considered a crossover-based local search (XLS) [35] as improvement method. This SSvariant also considers a restart mechanism to avoid local minima.

4. Coral Reefs Optimization

This section reviews the CRO-SL algorithm considered in this contribution for tackling the IR problem. First, we present the basic CRO algorithm, which will be modified with a *substrate layer* in order to obtain a competitive co-evolutionary algorithm with different exploration procedures involved.

4.1. Basic CRO

The CRO [14,36] is an evolutionary-type algorithm based on the behavior of the processes occurring in a coral reef. Let $\mathcal{R}$ be the reef represented by an $R_1 \times R_2$ grid, where each position (i,j) of $\mathcal{R}$ is able to allocate a coral or a colony of corals, $C_{i,j}$, standing for solutions to the current optimization problem at hand. The CRO algorithm first initializes some random positions of $\mathcal{R}$ with random corals, and leaves some other positions empty. These holes in the reef are available to host new corals that will be able to freely settle and grow in later phases of the algorithm. The rate between free/occupied positions in $\mathcal{R}$ at the beginning of the algorithm is a parameter of the CRO algorithm, denoted as ρ_0 with $0 < \rho_0 < 1$.

The second phase simulates the processes of reproduction and reef formation. The different reproduction mechanisms available in nature are recreated by sequentially applying different operators:

1. **External sexual reproduction or Broadcast Spawning**. Broadcast spawning consists of the following steps at each iteration k of the algorithm:
   1.a A random fraction of the existing corals is selected uniformly, turning these corals into broadcast spawners. The fraction of broadcast spawners with respect to the overall amount of existing corals in the reef will be denoted as F_b.
   1.b Several coral larvae are formed. To generate each new larva, two broadcast spawners are selected and a crossover operator or any other exploration strategy is applied. Note that once two corals have been selected to be the parents of a larva, they are not chosen anymore at iteration k for reproduction purposes. Corals' selection can be done randomly, uniformly, or using any fitness proportionate selection approach (e.g. roulette wheel).
2. **Internal sexual reproduction or Brooding**. Hermaphrodite corals mainly reproduce by brooding. This reproduction is modelled by means of any kind of mutation mechanism and takes place on a fraction of corals of $1 - F_b$. A percentage P_i of the coral is mutated.
3. **Larvae setting**. Once the larvae are formed either through external or internal reproduction, they will try to set and grow in the reef. Each larva will randomly try to set in a position (i,j) of the reef and, if the location is free, it will set. If the location is already occupied, the new larva will set only if its health function (fitness) is better than that of the existing coral. Moreover, the CRO algorithm defines a parameter η that determines the maximum number of tries a larva can attempt to occupy a position at each iteration k.
4. **Asexual reproduction**. Corals reproduce asexually by budding or fragmentation. CRO models this mechanism in the following way: the whole set of corals in the reef are sorted according to their level of health value (given by $f(C_{ij})$). Then, a small fraction (denoted as F_a) of the available corals are duplicated and mutated (with probability P_a) to provide variability, and try to settle in a different part of the reef as in Step 3.
5. **Depredation**. Corals may die during the reef's formation. Therefore, at the end of each reproduction iteration k, a small number of corals in the reef can be depredated, thus liberating space in the reef for next coral generation (iteration $k+1$). The depredation operator is applied with a very small probability (P_d) to a fraction (F_d) of the corals in the reef with worse health.

Algorithm 1 illustrates the flowchart diagram of the CRO algorithm, with the different CRO phases (reef initialization and reef formation), along with all the operators described above.

Algorithm 1 Pseudo-code for the CRO algorithm
Require: Valid values for the parameters controlling the CRO algorithm
Ensure: A single feasible individual with optimal value of its *fitness*
1: Initialize the algorithm
2: **for** each iteration of the simulation **do**
3: Update values of influential variables:
predation probability, etc.
4: Sexual reproduction processes
(broadcast spawning and brooding)
5: Settlement of new corals
6: Predation process
7: Evaluate the new population in the coral reef
8: **end for**
9: Return the best individual (final solution) from the reef

4.2. CRO with substrate layers (CRO-SL)

The original CRO algorithm described above is based on the main processes of coral reproduction and reef formation that occur in nature. However, there are many more interactions in real reef ecosystems that can be also modelled and incorporated into the CRO approach to improve its performance in optimization and search problems. For example, different recent studies have shown that successful recruitment in coral reefs (i.e., successful settlement and subsequent survival of larvae) strongly depends on the type of substrate on which they fall after the reproduction process [37]. This specific characteristic of coral reefs was first included in the CRO algorithm in Ref. [15] in order to solve different instances of the Model Type Selection Problem for energy applications. This new version of the CRO was named CRO-SL (Coral Reefs Optimization algorithm with Substrate Layers). In Ref. [38], a novel version of the CRO was presented, where several substrate layers which provide a different search procedure each were defined in the CRO process. Following this version of the algorithm, the CRO-SL general approach aims for competitive co-evolution, where each substrate layer represents different processes (different models, operators, parameters, constraints, repairing functions, etc.). This idea of CRO with substrate layers (CRO-SL hereafter) as a competitive co-evolution algorithm was successfully tested in different applications and problems such as micro-grid design [39], vibration cancellation in buildings [40], and in the evaluation of novel non-linear search procedures [41]. In this section we describe the main ideas of the CRO-LS as a co-evolutionary search algorithm, which will be applied to solve the problem of intensity-based medical IR faced in this contribution.

The inclusion of substrate layers in the CRO can be done in a straightforward way: we redefine the artificial reef considered in the CRO in such a way that each cell of the square grid Ψ representing the

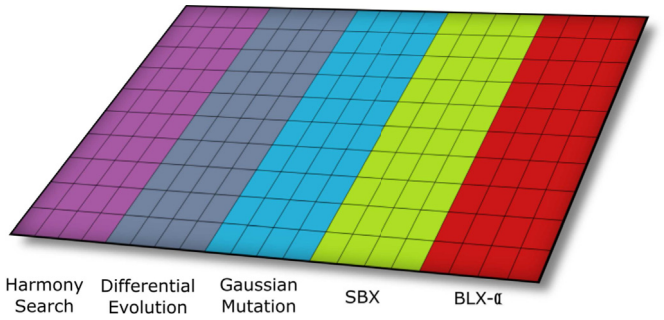

Fig. 3. Example of different substrate layers in a CRO-SL algorithm.

reef is now defined by 3 indexes (i, j, t), where i and j stand for the cell location in the grid, and index $t \in T$ defines each substrate layer by indicating which structure (search operator in this case) is associated with the cell (i, j) of the reef. Each coral in the reef is then processed in a different way (with a different search operator) depending on the specific substrate layer in which it falls after the reproduction process. The modification of the basic algorithm does not imply any change in the corals' encoding. Fig. 3 shows an example of the CRO-SL reefs structure, with five different substrate layers. In this example, each substrate is assigned to a different exploration process: Harmony Search, Differential Evolution, Gaussian mutation, SBX, and BLX-α.

5. CRO-SL for medical image registration

In this section we introduce the particular characteristics of the proposed algorithm. Following the reef analogy, a coral encodes a candidate solution to the optimization problem. In our specific IR approach, this solution is encoded as a real-valued vector that represents the parameters of a transformation model. The quality of a solution is evaluated using a fitness function. CRO-SL has been designed following both feature- and intensity-based approaches in a way that any similarity metric can be used directly as fitness function. Thereby, our proposal is able to accommodate different transformation models and similarity metrics providing a great flexibility when designing the experimental setup. The considerations regarding transformation models and similarity metrics for the experimental setup are detailed in Section 5.1. The specific design decisions related to the CRO and CRO-SL methods are presented in Section 5.2.

5.1. Coding scheme, fitness function and specific aspects related to the IR approach

Regarding the implementation of the CRO-based algorithms, our intention was to maximize the flexibility and provide a modular design to easily adapt our methods to different problem scenarios. To that aim, the scheme of the evolutionary algorithm is independent of the specific registration framework employed. On the one hand, any real-coded operator can be integrated as a substrate layer for the CRO-SL method. On the other hand, both the coding scheme and the fitness function can be replaced, so any transformation model or similarity metric can be considered.

In particular, the multi-modal experimental study addressed in this contribution considers an affine transformation model. Thus, solutions were encoded as twelve-dimensional real vectors: rotation versor $(\theta_x, \theta_y, \theta_z)$, shear factor (g_x, g_y, g_z), translation vector (t_x, t_y, t_z), and scaling factor (s_x, s_y, s_z). The search space was also restricted, considering $[-50, 50]$ as the parameters' range for the translation component, $[0.9, 1.1]$ for the scaling factor, a rotation angle in the $[-90, 90]$ range, and shearing in the range of $[-20, 20]$.

Next, we detail the specific design decisions for the different approaches:

5.1.1. Intensity-based approach

This approach was implemented in Elastix [58], a free, open-source, and widely-used toolbox specifically designed for intensity-based medical IR. Elastix is based on the *Insight Segmentation and Registration Toolkit* (ITK), which provides a thoroughly tested environment. The optimizer of intensity-based approaches is guided by the normalized mutual information (NMI) [42] metric. NMI was specifically proposed for 3D medical IR, due to being less sensitive to large misalignment between images, and is computed as:

$$\mathrm{NMI}(I_A, I_B) = \frac{\sum_{a\in I_A}\sum_{b\in I_B} p_{AB}(a,b)\log(p_A(a)p_B(b))}{\sum_{a\in I_A}\sum_{b\in I_B} p_{AB}(a,b)\log p_{AB}(a,b)},$$

where p_{AB} is the joint probability and p_A, p_B are the marginal discrete probabilities of the intensity values of the images.

In addition, the design of the intensity-based approach integrates two specific components: i) A **multi-resolution strategy**. In order to reduce the computational cost of the registration, the input images are pre-processed before the registration, applying both down-sampling and Gaussian smoothing to create two image representations, called pyramids. In the first resolution, the optimizer selects a small, low-detail pyramid of the input images, which allows the algorithm to obtain an approximation of the desired transformation. Thus, the similarity metric is computed over a small amount of random spatial samples (10,000). The second resolution is meant to be a refinement phase, improving the quality of the best transformations found during the first resolution. In this case, the number of spatial samples is increased up to 20,000. ii) A **restart mechanism**. The main objective of the restart mechanism is to ensure the process can recover from stagnation and find a good final solution. During the first resolution, the whole sequence of the optimization procedure is performed a fixed number of times, restarting the population at the end of each run and storing the best solution found. This solution is carried out to the second resolution, where the refinement is performed. As the first resolution is using a reduced version of the data, this mechanism is computationally cheap. A basic CRO method was proposed in Ref. [43] following this design and achieving promising preliminary results.

5.1.2. Feature-based approach

It was designed following a similar structure to the intensity-based method and implemented in C++. The optimizer is guided by the MedSE metric, which is robust in the presence of outliers and can be formulated as: $F(f, I_s, I_m) = w_1 \cdot (1/(1 + \sum_{i=1}^{N} \|(sR\vec{p_i} + \vec{t_i}) - \vec{p_j}'\|)) + w_2 \cdot (1/(1 + |p_c^s - p^m|))$, where I_s and I_m are the scene and model images; f is the solution encoding the transformation parameters; p_i is the i-th 3D point from the scene and p_j its corresponding closest point in the model obtained with a grid closest point data structure [34]; w_1 and w_2 (w_1+$w_2 = 1$) weigh the importance of each function term; p_c^s is the radius of the sphere wrapping up the transformed scene image; and p^m the radius of the sphere wrapping up the model image. Note that the first term of F corresponds to the MedSE between neighbor features.

Regarding specific components integrated within the original scheme of the feature-based algorithms, a restart mechanism is applied when the optimizer detects stagnation in the population. As feature-based approaches deal with a small set of feature points, no multi-resolution strategy is required.

5.2. Coral Reef Optimization methods design issues

In the case of the CRO-SL tested in this paper, each substrate layer only affects to the calculation of the larvae coming from the broadcast spawning process, whereas we have considered the same brooding procedure for all the corals in the reef. There are some important remarks

that can be done regarding the CRO-SL approach. First, note that the original CRO is a meta-heuristic based on exploitation of solutions, and leaves the specific exploration open (in the same manner as, for example, Simulated Annealing). This way, the CRO-SL can be seen as a generalization of the original CRO that does not modify the dynamics of the algorithm, with can still follow the basic steps given in the previous subsection. The only difference between the CRO-SL and the basic CRO is the specific implementation of the broadcast spawning procedure, which now depends on the specific substrate to which each coral (solution) is associated. Second, note that the CRO-SL is in fact a competitive co-evolutionary search procedure: the substrate layers of the CRO-LS promote the co-evolutionary process between different search operators, without the necessity of defining different populations. The competitiveness of the approach is given by the fact that CRO dynamics is based on a procedure of larvae settlement which directly involves competition among corals.

The considered substrates are detailed below. Note that there are general purpose substrates, such as Differential Evolution or Harmony Search-based, and other very specific substrates with crossovers adapted to the problem at hand, intensity-based medical IR.

- **Differential Evolution-based operator** (DE): This operator is based on the evolutionary algorithm with that name [44], a method with powerful global search capabilities. DE introduces a differential mechanism for exploring the search space. Hence, new larvae are generated by perturbing the population members using vector differences of individuals. Perturbations are introduced by applying the rule $v_i = x_i^1 + F(x_i^2 - x_i^3)$ for each encoded parameter on a random basis, where v corresponds to the output larva, x^t are the considered parents (chosen uniformly among the population), and F determines the evolution factor weighting the perturbation amplitude.
- **Harmony Search-based operator** (HS): Harmony Search [45] is a population based MH that mimics the improvisation of a music orchestra while its composing a melody. This method integrates concepts such as harmony aesthetics or note pitch as an analogy for the optimization process, resulting in a good exploratory algorithm. This property is also applicable to the operator based on its behavior. HS controls how new larvae are generated in one of the following ways: i) with a probability HMCR $\in [0,1]$ (Harmony Memory Considering Rate), the value of a component of the new larva is drawn uniformly from the same values of the component in the other corals. ii) with a probability PAR $\in [0,1]$ (Pitch Adjusting Rate), subtle adjustments are applied to the values of the current larva, replaced with any of its neighboring values (upper or lower, with equal probability).
- **BLX-α crossover** (BLX): BLX-α [29] is considered a standard recombination operator in IR due to its good performance and its ability to establish a solid exploration–exploitation trade-off. Two parents from the reef population are provided as input. For each position i of their encoding, the difference $d_i = |x_i - y_i|$ is computed and the output value v_i is randomly generated in the interval $[\min(x_i, y_i) - \alpha d, \max(x_i, y_i) + \alpha d]$, where α restricts the width of the range. Even though BLX-α generates two offspring by applying twice the operator, only the best one becomes the larvae that gets to compete for a space in the reef.
- **Simulated binary crossover** (SBX): SBX [46] is another recombination operator commonly used in real-coded optimization. It attempts to simulate the offspring distribution of the binary-coded one point crossover. In SBX, two new larvae v_i^1 and v_i^2 are created in proportion to the difference between two parent solutions, x_i^1 and x_i^2. The first step is to define a spread factor as: $\beta_i = |(v_i^2 - v_i^1)/(x_i^2 - x_i^1)|$. Next, from a specified probability distribution function:

$$P(\beta_i) = \begin{cases} 0.5(\eta+1)\beta_i^{\eta}, & \text{si } \beta_i \le 1 \\ 0.5(\eta+1)/\beta_i^{\eta+2}, & \textit{otherwise} \end{cases}$$

the ordinate β_q is found so that the area under the probability curve from 0 to β_q is equal to a random number u:

$$\beta_q = \begin{cases} (1/(2(1-u)))^{\frac{1}{\eta+1}}, & \text{si } u \le 0.5 \\ (2u)^{\frac{1}{\eta+1}}, & \textit{otherwise} \end{cases}$$

η corresponds to the distribution index, being any non-negative real number. A large value of η provides a large probability for creating near parent solutions while a small value produce distant offspring solutions. Finally, the descendant solutions are calculated as:

$$v_i^1 = 0.5[(1+\beta_q)x_i^1 + (1-\beta_q)x_i^2],$$
$$v_i^2 = 0.5[(1-\beta_q)x_i^1 + (1+\beta_q)x_i^2].$$

In the same way as BLX-α, only the best offspring prevails and reaches the larvae settlement phase.

- **Gaussian Mutation** (GM): This operator is a variant of the classical random mutation method introducing a scaled Gaussian distribution [47]. Specifically, the Gaussian probability density function is:

$$f_{G(0,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

This operator has a mutation strength value σ_i for every parameter of a solution, related to the lower and upper bounds a_i and b_i, respectively. Thus, $\sigma = \sigma_i/(b_i - a_i)$ is used as a fixed non-dimensionalized step-size whose value also depends on time, as it decreases based on the number of the current iteration. The reason of adapting the value of σ is to provide a stronger mutation at the beginning of the optimization, while fine tuning with smaller displacements nearing the end. The mutated larva is thus calculated as: $v_i = x_i + \sigma N_i(0,1)$, where $N_i(0,1)$ is a random number following the Gaussian distribution.

- **Cauchy Mutation** (CM): If the search space is distributed in such way that relative minima are far apart, the mutations provided by GM may lead to premature convergence to a local minima, even considering an adaptable σ. In these cases, the Cauchy operator [48] might be suitable, due to the wider scope of its probability distribution. On the other side, this distribution provides small mutations with a reduced probability, leading to a poor accuracy in the local search. The Cauchy density function is: $f_{C(0,t)}(x) = \frac{t}{\pi(t^2 + x^2)}$, where $t > 0$ is a scale parameter. The offspring generated by CM is calculated as: $v_i = x_i + \sigma C_i(0,1)$, where $C_i(0,1)$ is a random number following the Cauchy distribution.
- **Strange Attractors** (SA): SA is a search operator proposed in Ref. [41], specifically designed to improve the searching capabilities of MHs by replicating different processes in nature known as fractal geometric patterns [49]. In particular, it is designed to generate structures of non-linear dynamical systems with chaotic behavior [50]. Overall, these kinds of fractal structures can be generated by means of the general two-dimensional quadratic map:

$$\begin{cases} x_{n+1} = a_1 + a_2 x_n + a_3 x_n^2 + a_4 x_n y_n + a_5 y_n + a_6 y_n^2 \\ y_{n+1} = a_7 + a_8 x_n + a_9 x_n^2 + a_{10} x_n y_n + a_{11} y_n + a_{12} y_n^2 \end{cases}$$

This quadratic map allows us to generate very different chaotic-behavior attractors with a reduced number of input parameters. In order to introduce a chaotic mutation in a solution using SA, we followed the procedure introduced in Ref. [41]. Hence, a determined number of attractors $S = 50$ are defined. Each time the operator is applied, a random attractor is considered. The quadratic map is calculated over a random number $\epsilon \in [500, 1000]$ of iterations starting from an initial condition $(x_0, y_0) = (0.6, 0.9)$ until x and y are found. Then, the mutated

larvae is obtained as:

$$\begin{cases} v_i = x_i + Fx; & p = 1/2 \\ or \\ v_i = x_i + Fy; & p = 1/2 \end{cases}$$

where F controls the mutation width, and p is the probability of choosing x or y. Fig. 4 shows some examples of strange attractors, regarding a simulation with values of parameters a_1–a_{12} in the range [−1.2, 1.2]. This range permits the generation of an extremely large variety of attractors (over 25^{10}) with different characteristics: chaotic, intermittent or convergent to a periodic orbit.

5.3. Choice of the best CRO-SL configuration

We developed a specific experimentation to comprehensively analyze the behavior of the substrate layer proposal in comparison with other relevant evolutionary IR methods. The results of this preliminary experimental study are detailed in Appendix A. The best configuration for the CRO-SL method was determined by a thorough study on the performance of the previous operators in order to find a suitable combination of substrate layers (see Section Appendix A.3). In particular, the best CRO-SL scheme involves a well-balanced combination of two crossover operators (BLX-α and SBX), two evolutionary-based operators (HS and DE), and one mutation (GM).

6. Affine multi-modal experimental study

In this section, we consider CRO to deal with a real-world medical IR problem where the registration scenario entail a tougher challenge. We test the performance of the most robust evolutionary intensity-based IR algorithms from the preliminary experiment developed in Appendix A and compare their results with different classical approaches.

6.1. Image dataset and problem scenarios

The images were obtained from the Retrospective Image Registration Evaluation Project (RIRE) [17]. This dataset features many multimodal registration scenarios for up to eighteen different patients. We considered nine patients (P101-P109) and three of these modalities, CT to MR-PD (4 scenarios), CT to MR-T1 (9 scenarios), and CT to MR-T2 (8

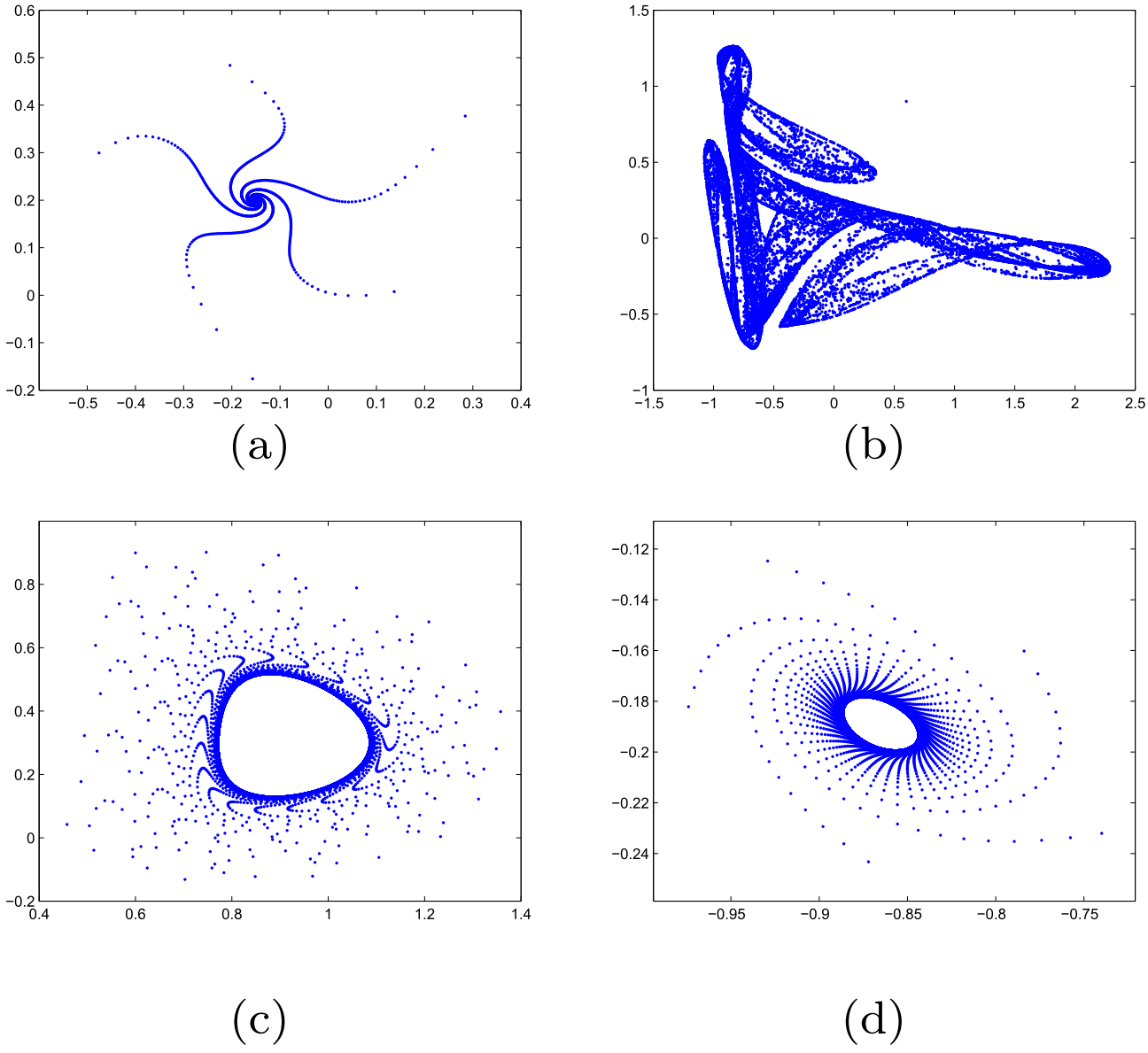

Fig. 4. Examples of four Strange Attractors in the phase space (x vs. y), used in the SA substrate. Each pattern corresponds to a different set of parameters (a_1-a_{12}).

Table 1
Detailed results for the registration error (mm) of each IR scenario in the RIRE dataset considering an affine transformation. Evolutionary methods include the averaged result of the 30 runs and the standard deviation. The last row shows the number of times an algorithm obtains the best outcome.

Scenario	Simplex	FLIRT	CRO-SL*		CRO*		SS*		ASGD	
			μ	sd	μ	sd	μ	sd	μ	sd
PD-P101	14.74	3.26	**3.13**	0.2	3.36	0.3	3.40	0.7	3.28	0.1
PD-P102	9.05	2.22	2.36	0.5	2.45	0.6	2.78	0.9	**2.18**	0.0
PD-P103	**3.03**	3.55	3.53	0.5	3.56	0.5	3.78	0.8	3.50	0.0
PD-P104	4.47	**2.34**	2.49	0.3	2.51	0.3	2.89	0.9	2.72	0.0
T1-P101	18.46	**2.20**	2.26	0.2	2.38	0.3	2.77	0.8	2.49	0.1
T1-P102	9.50	**1.54**	1.80	0.5	1.90	0.5	2.34	1.2	1.58	0.0
T1-P103	2.45	2.56	**2.43**	0.5	2.62	0.4	2.70	0.7	2.53	0.1
T1-P104	18.38	1.59	**1.36**	0.3	1.40	0.5	2.59	1.4	1.74	0.0
T1-P105	13.55	2.70	**2.47**	0.7	3.33	1.4	4.13	2.0	4.01	3.5
T1-P106	8.35	**2.09**	2.12	0.2	2.26	0.4	2.70	1.0	2.43	0.9
T1-P107	24.24	**1.15**	1.24	0.3	1.52	0.4	2.25	1.2	1.29	0.1
T1-P108	10.94	**1.77**	1.79	0.2	1.89	0.4	2.31	1.0	1.92	0.0
T1-P109	4.58	**1.41**	1.43	0.4	1.73	0.3	2.09	0.7	1.76	0.0
T2-P101	10.09	2.97	**2.77**	0.3	2.84	0.3	3.09	0.8	2.81	0.1
T2-P102	8.73	2.01	1.95	0.6	1.95	0.6	2.45	1.0	**1.74**	0.1
T2-P104	4.52	1.86	1.76	0.4	1.76	0.4	2.14	0.8	**1.71**	0.0
T2-P105	14.00	2.79	**2.52**	0.3	2.65	0.5	3.35	1.1	2.63	0.1
T2-P106	8.45	2.91	**2.55**	0.3	2.72	0.4	3.05	0.9	2.72	0.1
T2-P107	6.47	1.88	2.10	0.5	2.31	0.6	2.56	1.0	**1.86**	0.1
T2-P108	15.48	2.14	**1.85**	0.2	2.04	0.3	2.25	0.7	1.93	0.1
T2-P109	4.48	2.24	**1.82**	0.3	2.18	0.3	2.48	0.7	2.19	0.1
# wins	1	7	9		–		–		4	

The best averaged results for each scenario are highlighted in bold.

scenarios). CT images have a size of $512 \times 512 \times 28 - 34$ voxels while MR images have a size of $256 \times 256 \times 20 - 26$ voxels.

6.2. Experimental design and parameter configuration

We designed an initial tuning phase to establish a common framework for the comparison of the different methods. This tuning was performed considering three scenarios from a training set available at the RIRE website, which includes a ground truth to validate the initial results. Once a common experimental framework for all methods was established, we performed the experiments over the test scenarios. The registration error was computed internally by the RIRE system as the average distance of the corner points of different volumes of interest (VOIs).[1]

We compared the best performing evolutionary algorithms (SS*, CRO*, and CRO-SL*) against three classical methods. On the one hand, we considered the Elastix implementations of two well-known optimization algorithms: Nelder-Mead or downhill simplex (Simplex) [24] and Adaptive Stochastic Gradient Descent (ASGD) [6]. On the other hand, we also included another state-of-the-art method named FLIRT [5], which stands out in the literature as a robust and accurate IR method.

As previously mentioned, the considered scenarios involve a rigid registration of different image modalities of the same patient. Appendix B summarizes the experimental results of applying simple rigid body transformations. In order to increase the complexity of the optimization problem, we considered an affine transformation model for the scenarios in this experiment. Hence, the addressed problem is considerably hardened, shifting from 6 to 12 DOF in the parameter search space.

A common experimental framework was designed for those methods implemented in Elastix, using NMI as similarity metric, a multiresolution strategy (computing the metric over 10000 and 20000 random samples), and 280 s of time limit (Simplex converged after 30 s). Internal parameters of the evolutionary methods were maintained with respect to the preliminary experimental setup (the design particularities are detailed in Appendix A.2). FLIRT, however, was executed in the framework provided by the registration library FSL [25], also considering NMI as similarity metric but employing a particular global optimization strategy [5]. On average, the registration process of FLIRT takes approximately 300 s per scenario. Thirty runs were performed per instance to provide an unbiased analysis for the stochastic methods. A computer with an Intel Core i7-4790 3.6 GHz processor and 16 GB RAM was used for the execution of the experiments.

[1] A detailed description of the evaluation process is available at www.insight-journal.org/rire/information.php.

Table 2
Ranking obtained through Friedman's test and statistical p-values with CRO-SL* as control method for Bonferroni's test, and Holm's test according to the averaged registration error.

	Rank	Bonferroni p	Holm p
CRO-SL*	1.86	–	–
FLIRT	2.57	1	0.28
ASGD	2.71	0.69	0.28
CRO*	3.19	0.10	0.06
SS*	5.10	$< 10^{-07}$	$< 10^{-08}$
Simplex	5.57	$< 10^{-10}$	$< 10^{-10}$

6.3. Analysis of experimental results

Here we perform an extensive comparison of the performance of the proposed CRO-based methods and relevant classical IR methods. For each method and scenario, the standardized RIRE evaluation system provides the registration error of ten VOIs. Table 1 reports the averaged results of these ten volumes for each IR method regarding twenty one different scenarios. Non-deterministic methods include the average error and the standard deviation of the thirty runs. In addition, we complement the analysis including the average ranking of the algorithms and apply different post-hoc statistical analyses, which will allow us to detect significant differences among the algorithms if they exist [51]. In particular, we consider Friedman's nonparametric test [52], Bonferroni-Dunn's test [53], and Holm's test [54] (Table 2). Last, we assess the quality of the results by visually evaluating the images obtained from the registration process.

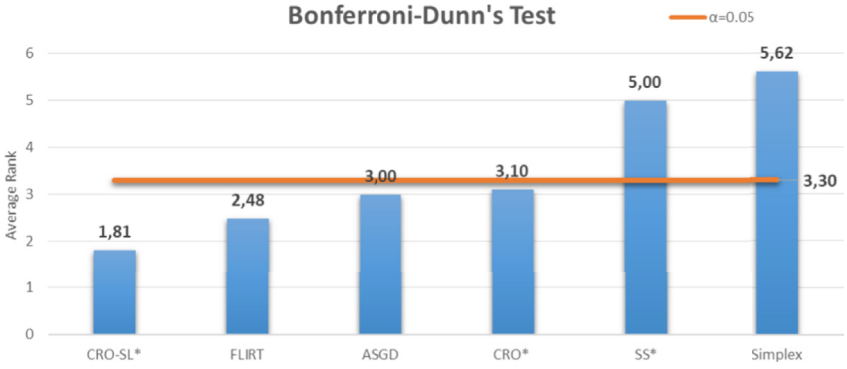

Fig. 5. Bonferroni-Dunn's graphic corresponding to the averaged test results for the multi-modal experimentation.

Among classical methods, Simplex delivered the most unstable results, ranking in the last position, failing to converge (reaching errors up to 24 *mm*) in eight instances. Even so, Simplex was able to obtain the minimum error in one occasion (ranking in last position). ASGD showed a standard but robust performance reaching a maximum error of 4.01 *mm* and achieving the minimum result in four scenarios (ranking third) with low standard deviation values. FLIRT stands out by behaving consistently among all the scenarios, reaching the minimum error in seven instances. Nevertheless, FLIRT ranked second (Table 2) with a score of 2.57, being the best classical method in this comparison.

Regarding evolutionary methods, SS* results show a regular performance, reaching a maximum error of 4.13 *mm*. Nonetheless, SS* ranked in the fifth place. CRO* behaved significantly better than SS*, displaying a lower variability, and reaching a medium ranking score (3.19). Neither SS* nor CRO* were able to obtain the minimum error in any particular scenario. In turn, CRO-SL*was the evolutionary approach with lowest standard deviation values, outperforming both evolutionary methods and achieving the minimum result in nine scenarios. CRO-SL* scored the best ranking position (1.86).

Regarding the statistical analysis, we first considered Friedman's nonparametric test [52], which aims to test the null hypothesis stating that there is no difference among the resulting mean errors of all the algorithms. The result of applying Friedman's test is $\chi_F^2 = 66.65$, and the corresponding p-value is $< 10^{-13}$. Given that the p-value is lower than the considered level of significance ($\alpha = 0.01$), the test concludes that there are significant differences among the results.

We also considered the Bonferroni-Dunn test [53] to detect significant differences among a control approach and the rest of the algorithms. In this case, the control algorithm is the one with the best ranking, CRO-SL*. The results are represented in Fig. 5, which illustrates the significance of the difference among rankings obtained for each algorithm considering the mean value of the quality metric. Vertical bars correspond to the sum of the ranking of the control method CRO-SL* and the critical difference value computed by the Bonferroni method. The algorithms exceeding the horizontal line (significance threshold) perform significantly worse than the control method for the considered $\alpha = 0.05$. We complemented the analysis with Holm's test comparing CRO-SL* with the rest of the methods. The p-value results for the

Table 3
Detailed results for the registration error (mm) of each IR scenario in the RIRE dataset. For each scenario, the comparison features the minimum error of the ten VOIs, the overall average error, and the averaged error of the best run of CRO-SL*. The last row shows the number of times an algorithm obtains the best outcome.

Scenario	Average error			minimum error	
	FLIRT	CRO-SL*		FLIRT	CRO-SL*
		Mean (all runs)	Best run		Mean (all runs)
PD-P101	3.26	**3.13**	2.87	3.21	**2.96**
PD-P102	**2.22**	2.36	2.02	**1.51**	1.55
PD-P103	3.55	**3.53**	3.28	**2.41**	2.51
PD-P104	**2.34**	2.49	2.24	**1.81**	1.85
T1-P101	**2.20**	2.26	1.99	**2.04**	2.06
T1-P102	**1.54**	1.80	1.44	**0.83**	1.03
T1-P103	2.56	**2.43**	2.14	1.79	**1.48**
T1-P104	1.59	**1.36**	0.86	1.30	**1.10**
T1-P105	2.70	**2.47**	2.12	1.56	**1.37**
T1-P106	**2.09**	2.12	1.93	**1.80**	1.85
T1-P107	**1.15**	1.24	0.87	**0.53**	0.75
T1-P108	**1.77**	1.79	1.53	1.71	**1.64**
T1-P109	**1.41**	1.43	1.00	1.00	**0.70**
T2-P101	2.97	**2.77**	2.44	2.82	**2.55**
T2-P102	2.01	**1.95**	1.54	**1.21**	1.25
T2-P104	1.86	**1.76**	1.43	1.32	**0.97**
T2-P105	2.79	**2.52**	2.12	2.40	**1.94**
T2-P106	2.91	**2.55**	2.22	2.52	**2.18**
T2-P107	**1.88**	2.10	1.89	1.18	**1.06**
T2-P108	2.14	**1.85**	1.56	1.93	**1.61**
T2-P109	2.24	**1.82**	1.60	1.47	**1.23**
# wins	9	**12**		8	**13**
	1		**20**		

The minimum registration errors for each scenario are underlined, and the best averaged results are highlighted in bold.

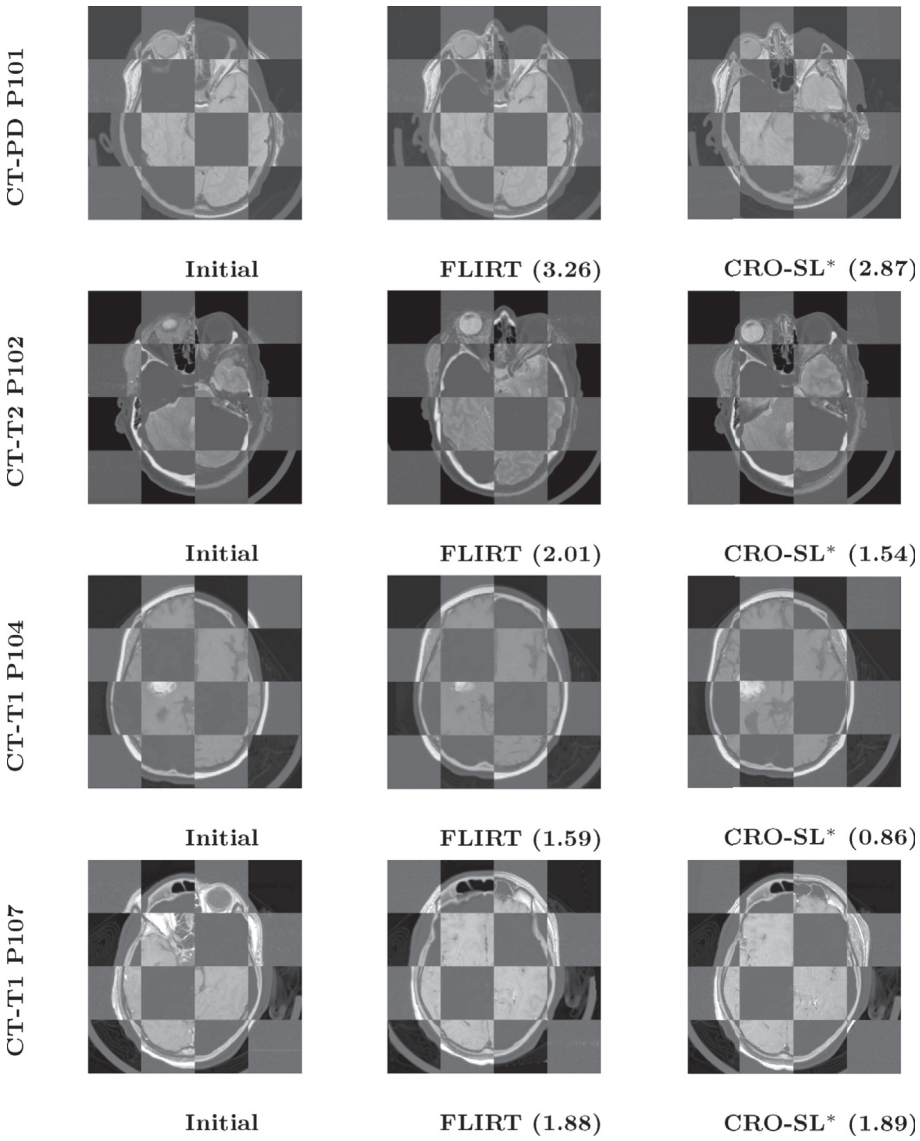

Fig. 6. Checkerboard visualization of the overlapping between multi-modal CT-MR scans. Figures provide a comparison of the results obtained by FLIRT and CRO-SL* in four scenarios PD-P101, T2-P102, T1-P104, and T1-P107.

remaining statistical tests reveal significant differences with SS* and Simplex (Table 2).

In view of these results, there are no definite differences between classical and evolutionary approaches, and no method clearly stands out in the comparison. Therefore, we included an in-depth analysis of the best performing methods with the aim to assess the behavior of our proposal, considering FLIRT as the best classical method and CRO-SL* as the best evolutionary approach. Table 3 presents the results of the minimum error of the ten VOIs, the overall average error per scenario, and the average error concerning the best of the 30 runs achieved by CRO-SL*.

In a direct comparison between FLIRT and CRO-SL*, the differences between both approaches become noticeable. When considering averaged results, CRO-SL* outperformed FLIRT in 12 scenarios. FLIRT achieved the best outcome in the remaining 9 scenarios. According to the averaged registration error per scenario, CRO-SL* provides slightly more accurate results with errors ranging from 0.24 to 4.71 *mm*, while FLIRT errors lie in the interval 0.53 to 3.99 *mm*. Furthermore, concerning the minimum VOI error per scenario our proposal stands out by achieving the most accurate results in 13 of the 21 instances. Finally, considering the best run of CRO-SL* the differences are remarkable. Our proposal is able to deliver the most accurate results with low standard deviation values (an average of 0.15 *mm* among runs). Thus, the best execution of CRO-SL* outperforms the best classical method in 20 of the 21 scenarios.

Summarizing, CRO-SL* is able to outperform both classical and evolutionary state-of-the-art IR methods of the multi-modal comparison. Even though there are no significant differences between CRO-SL* and FLIRT when comparing averaged results, our proposal proved to be a robust and efficient approach, capable of delivering more accurate results than an state-of-the-art method in affine IR.

Finally, we include a visual assessment of the best performing methods for two registration scenarios. Fig. 6 presents a visualization of the overlapping between multi-modal images regarding the solutions obtained by FLIRT and CRO-SL*. No substantial differences can be appreciated when analyzing the outcome of the registrations, as both approaches deliver accurate outcomes.

7. Conclusions

In this work, we described the design and implementation of a novel nature-inspired technique, known as CRO-SL algorithm, to solve the problem of 3D medical IR in an efficient and robust way. CRO-SL is an enhanced version of an algorithm based on the natural processes taking place within a coral reef, such as reef formation and expansion, coral reproduction, and fight for a space in the colony. This algorithm

integrates a powerful and competitive co-evolutionary strategy in its design, which has allowed it to obtain excellent results in different complex real-world problems. The use of different exploration strategies within the evolutionary procedure provides strong indications for a precise, robust and efficient optimization algorithm. Hence, we adapted and applied CRO-SL to the IR problem following both intensity- and feature-based approaches to analyze its behavior. The structure of our proposal was designed in such way that any transformation can be directly used. For example, in case a deformation model is required, e.g. a non-rigid B-spline transformation, the only necessary modification affects the codification of the transformation. This change is trivial in Elastix, which only involves the order of the b-spline as an additional parameter. The internal architecture of the operators integrated within the CRO-SL scheme is not affected in any way.

Continuous optimization methods perform an exhaustive local search in the space of the transformation parameters, an approach sensitive to the initial conditions and prone to get trapped in local minima. On the contrary, discrete optimization methods perform a global search, robust to the initial conditions and showing a better convergence rate and computational efficiency at the expense of the quantization of the search space. When compared to techniques such as discrete optimization that aim to overcome the disadvantages of continuous approaches, our proposal employs advanced optimization strategies. While the discretization of the search space comprises an implicit loss of precision, the use of a metaheuristic strategy allows us to guide the search in a continuous optimization space. Our proposal is thus an efficient alternative to exhaustive techniques without the precision loss of discrete approaches, based on the advantages of evolutionary computation: global search capability, robust and reliable performance, implementation flexibility, and limited or not required expert knowledge.

In order to evaluate the performance of our proposal, we compared both implementations of CRO-SL with some of the best performing IR methods in the literature. We tackled two broad IR experimental studies. First, we considered a preliminary experimental setup (see Appendix A) involving sixteen mono-modal IR scenarios obtained by pairing four different synthetic brain MRI images, with different levels of complexity regarding noise and sclerosis lesions. These images were extracted from the BrainWeb database at McGill University [16]. Next, we tested our proposal in a challenging real-world problem involving twenty one multi-modal IR scenarios. These images were provided by the RIRE project [17]. Through an extensive experimentation, CRO-SL provided outstanding results considering both registration approaches, outperforming classical and well-consolidated methods and demonstrating the efficiency, robustness and suitability of the proposed algorithm when tackling complex optimization problems such as medical IR.

Acknowledgments

This work has been partially supported by the projects TIN2015-67661-P, including European Regional Development Funds (ERDF), from the Spanish Ministery of Economy, and TIN2014-54583-C2-2-R from the Spanish Ministerial Commission of Science and Technology (MICYT).

Appendix A. Mono-modal experimental study

In this appendix we introduce a complementary experimental study to test the performance of our CRO-based proposals. The mono-modal experimentation was designed as a preliminary study to compare our proposal with the evolutionary IR methods described in Section 3. These methods are considered the state-of-the-art evolutionary algorithms for both feature- and intensity-based medical IR approaches. Thus, we aim to provide an exhaustive analysis of the results, considering optimization capabilities of the methods tackling a synthetic medical IR problem.

Appendix A.1. Image dataset and problem scenarios

The images used in this experiment were obtained from the BrainWeb public repository at McGill University [16]. This repository consists of a simulated brain database providing synthetic MRIs computationally generated. The BrainWeb repository has been widely used by the IR research community [55]. In order to consider different degrees of complexity, some images include noise and multiple sclerosis lesions, as detailed in Table A.4. All the images have the same size ($60 \times 181 \times 217$ voxels).

Table A.4
Characteristics of the brain MRI images used in our experimental study, including Gaussian noise level, presence of lesion, and number of extracted features.

Image	Lesion	Noise	Crest-line points
I_1	No	None	583
I_2	No	1%	393
I_3	Yes	1%	348
I_4	Yes	5%	284

For each of the original images we extracted a set of points containing relevant curvature information. These set of points were obtained using a 3D crest-line edge detector [56]. In feature-based approaches, the heuristic values of the crest-line points will guide the optimizer to estimate the registration transformation. In order to create different IR problem scenarios, each image was transformed using one of the four similarity transformations (involving rotation, translation, and uniform scaling) shown in Table A.5. Note that these parameter values of the transformations produce strong changes in the object location, generating complex IR scenarios.

Thus, we designed a preliminary test aimed at tuning the substrate layer design of the CRO-SL method. Specifically, this test involves four different IR instances: I_3 versus $T_i(I_4)$ for $i = 1, 2, 3, 4$. To avoid any bias during the experimentation, we discarded the tuning scenario for the comparison with the remaining evolutionary methods and created a distinct experimental scenario. Hence, the mono-modal experimental study features a total number of sixteen IR problem instances that were created by pairing images with different transformations. From lower to higher complexity, the scenarios are I_1 versus $T_i(I_2)$, I_1 versus $T_i(I_3)$, I_1 versus $T_i(I_4)$, and I_2 versus $T_i(I_4)$, for $i = 1, 2, 3, 4$.

Table A.5
Parameters of the transformations used in our experiments: rotation angle (λ), rotation axis (a_x, a_y, a_z), translation vector (t_x, t_y, t_z), and uniform scaling factor s.

	λ	a_x	a_y	a_z	t_x	t_y	t_z	s
T_1	115	−0.863	0.259	0.431	−26	15.5	−4.6	1
T_2	168	0.676	−0.290	0.676	6	5.5	−4.6	0.8
T_3	235	−0.303	−0.808	0.505	16	−5.5	−4.6	1
T_4	276.9	−0.872	0.436	−0.218	−12	5.5	−24.6	1.2

Appendix A.2. Experimental design and parameter configuration

We aim to provide an objective evaluation of the performance of many IR methods following different approaches. As reviewed in Section 2, these algorithms differ in their optimization procedure: feature-based approaches are typically guided by MSE or MedSE and consider only feature points, while intensity-based approaches are guided by either MI or NC considering the intensity values of the images. The resulting values of different metrics are not directly comparable and, furthermore, similarity metrics are not suitable for evaluating the quality of the registration. An alternative, as suggested in Ref. [57], is to consider an evaluation metric measured as a distance in space regarding identifiable anatomical locations. Therefore, a standardized framework is required in order to objectively evaluate the registration results. We considered a common measure to evaluate the quality of all solutions. Once the algorithm reaches a solution, the MSE distance between the transformed scene's features and the model's features is computed over a set of anatomical landmarks (crest-line points). This value will be used as final quality comparison for all methods without interfering with the particularities of each IR approach. In addition, we complement the analysis with a visual inspection of the resulting registration to assess the accuracy of the methods.

It is important to remark that we are comparing the MSE between crest-line points extracted from different images. Therefore, there is an implicit registration error in any of the previous IR problems. An IR method may estimate the ground-truth transformation that corresponds to the best possible alignment of the images (see Table A.5). However, the MSE of such a perfect IR method will not be null. After the perfect alignment of the images, there will remain some distances between the corresponding crest-line points of both images. Such distances are due to the effect of noise and/or lesions in the crest-lines extracted from the images. Table A.6 reports the minimum MSE (in mm) that can be achieved by any IR method when dealing with the IR problems presented in Section Appendix A.1.

Table A.6
Minimum MSE (in mm) achievable in scenarios involving paired images.

Scene Model	i_1 i_2	i_1 i_3	i_1 i_4	i_2 i_4	i_3 i_4
MSE	31.1	42.4	46.0	28.2	34.9

Regarding the coding-scheme of the mono-modal study, we considered a similarity transformation to relate the images, with $[-30, 30]$ as the parameters' range for the translation component, and $[0.75, 1.25]$ for the scaling factor. No restriction to the rotation axis was applied so the actual range for each component of the rotation versor is $[-1, 1]$, with a rotation angle in the $[0, 360]$ range. Thereby, the transformation is encoded using seven real-coded parameter solution if the approach is intensity-based: rotation versor $(\theta_x, \theta_y, \theta_z)$, translation vector (t_x, t_y, t_z), and uniform scaling factor s. In case of feature-based approaches, instead of using a versor, the rotation was encoded using three parameters for the rotation axis in the interval $[-1, 1]$, and a rotation angle in the $[0, 360]$ range. Thus, the solutions use eight-dimensional real coded transformations.

In order to design our experimental setup as comprehensively as possible, we adapted our CRO-based proposals (see Section 5) and the state-of-the-art IR methods (see Section 3) to both feature- and intensity-based approaches. We also considered a Multi-Start (MS) heuristic to provide a reference (baseline) for the performance of the different approaches. MS follows a simple optimization design, where a random solution is first generated and then improved using a local search. Hence, we will compare 10 different IR methods:

- **Intensity-based methods**[2]**:** MS*, r-GA*, SS*, CRO*, and CRO-SL*.
- **Feature-based methods:** MS, HE-GA, SS, CRO, and CRO-SL.

Next, we detail the distinct characteristics of the considered approaches:

Intensity-based methods design:

Every intensity-based method was implemented in Elastix and designed following the framework defined for our CRO-based proposals (see Section 5.1), i.e., seven-dimensional real-coded solutions, NMI as similarity metric, multi-resolution strategy, and restart mechanism. Due to the substantial amount of data that intensity-based approaches consider, the running time of each algorithm was limited at 180 s, including both resolutions.

As for the parameter configuration of each individual method (see Table A.7), we followed the guidelines proposed by their original authors and previous studies in IR [13,23]. Some of the parameters were manually adjusted for this particular experiment (i.e., population size, number of restarts, and number of generations, among others) to fulfill the running time limitations. We considered a 'parent-centric' crossover-based local search, called PMX-α [32], for those algorithms involving a solution improvement component.

[2] For the sake of clarity, and to differentiate between approaches, we noted every intensity-based method including an asterisk (*) after its name.

Table 7
Parameter configuration for the intensity-based IR methods.

r-GA*		SS*		CRO*		CRO-SL*	
individuals	100	Psize	12	reef size	80	reef size	65
generations	75	gen.	18	gen.	65	gen.	63
restarts	5	restarts	3	restarts	8	restarts	6
tournament size	3	PMX iters.	12	PMX	15	PMX	12
blend factor α	0.3	blend factor α	0.3	ρ_0	0.6	blend factor α	0.3
cross. prob.	0.5	reference set size	4	F_{broad}	0.8	HMCR	0.8
mut. prob.	0.1			P_d	0.15	PAR	0.2
				k	3	σ	0.45
						F_{de}	0.2

Feature-based methods design:

Feature-based methods were implemented in C++ and compiled with the GNU/g++ tool. Since MSE was chosen as quality metric for the final comparison, we considered a different optimization measure to avoid favoring feature-based methods. Hence, we adopted MedSE, which was specifically designed as similarity metric for feature-based medical IR approaches [59]. Still, we expect this choice to introduce a small bias.

Every proposal in this experimentation was implemented using eight real-coded parameter solutions and include a restart mechanism when the optimizer detects stagnation. The stopping criteria for every feature-based algorithm was limited to 20 s, due to the reduced sizes of the feature sets.

Table A.8 summarizes the algorithm configuration used in the feature-based approaches. The parameters of the different approaches were manually adjusted for this experiment, starting from the recommended values provided by the authors [39] or previous studies [12,60]. In this case, algorithms involving a local search make use of 100 iterations of XLS [35].

Table A.8
Parameter configuration for the feature-based IR methods.

HE-GA		SS		CRO		CRO-SL	
individuals	60	Psize	30	reef size	70	reef size	70
restarts	5	restarts	5	restarts	15	restarts	15
blend factor α	0.3	XLS iters.	100	XLS	100	XLS	100
cross. prob.	0.7	blend factor α	0.3	ρ_0	0.6	blend factor α	0.3
mut. prob.	0.2	reference set size	8	F_{broad}	0.9	HMCR	0.8
				P_d	0.05	PAR	0.2
				k	3	σ	1
						F_{de}	0.5

Appendix A.3. Substrate layer design for CRO-SL

In this section we perform a thorough preliminary analysis on the performance of the seven considered substrates for the proposed design of CRO-SL. We aim to identify a combination of operators with the best possible trade-off between exploration and exploitation. To do so, we incorporated a tuning scenario (I_3 versus $T_i(I_4)$) not included in the final experimentation, as it is only aimed for the adjustment of the CRO-SL components. This preliminary analysis will be solely focused on the behavior of the substrate layers attending to an optimization viewpoint. Therefore, we will single out the combination of substrates that better optimizes the considered NMI similarity metric together with the quality of the registration provided by the MSE. We performed the trial experiments considering the intensity-based approach implementation. Note that the main parameters of CRO-SL* are kept unchanged while the number of considered substrates is the only parameter adjusted.

First, we study the performance of configuring CRO-SL* using a single substrate and pairs of substrates (Table A.9). Some combinations suffer from early stagnation in many runs, as the operators involved in the design are not able to overcome local minima. On the contrary, there are configurations delivering both accurate and robust results, close to the minimum achievable MSE (34.9 *mm*). Attending to the NMI metric, the differences between most combinations of substrate layers are almost negligible. Nonetheless, many configurations deliver satisfactory results in terms of similarity metric yet fail to provide accurate registration results. As mentioned in Section Appendix A.2, the need of considering a standardized quality measure to evaluate the registration is thus undeniable. The preliminary results show that the BLX-α operator stands out among the seven considered operators, as it provides the best optimization results in most of the instances, including its combination with the rest of the operators.

Table A.9
Preliminary test of the low-level combinations of one and two substrates. Averaged results for the NMI and MSE similarity metrics between crest-line points are provided. In brackets, the minimum achievable MSE value (mm) is shown.

Substrates	NMI	MSE [35]	Substrates	NMI	MSE [35]
BLX	**−1.294**	**261.6**	GM	−1.262	475.7
SBX	−1.275	3879.4	CM	−1.237	3186.0
HS	−1.282	462.2	SA	−1.283	485.9
DE	−1.291	1097.1			
BLX SBX	−1.291	510.9	HS DE	−1.284	257.5
BLX HS	−1.287	1574.7	HS GM	−1.266	1090.2
BLX DE	−1.294	259.8	HS SA	−1.285	487.6
BLX GM	−1.292	257.5	HS CM	−1.256	328.8
BLX SA	−1.293	481.4	DE GM	−1.273	1835.2
BLX CM	**−1.294**	**36.62**	DE SA	−1.289	904.6
SBX HS	−1.284	2223.9	DE CM	−1.269	1413.1
SBX DE	−1.292	485.5	GM SA	−1.279	1324.2
SBX GM	−1.282	1558.1	GM SM	−1.257	1198.6
SBX SA	−1.288	1392.1	CM SA	−1.270	694.2
SBX CM	−1.280	2485.4			

The best substrate combinations are highlighted in bold.

Next, Table A.10 includes the results for medium-level combinations of three and four different operators. The results confirm the good performance provided by the two crossover operators, BLX-α and SBX. Combinations involving a crossover operator and HS or DE also obtain good performance in most of the occasions, while SA and both Gaussian and Cauchy mutations tend to affect negatively the performance of the algorithm. Those configurations where no crossover is applied perform considerably worse than the rest, due to the algorithm not being able to overcome a premature stagnation.

Table A.10
Averaged results for the NMI and MSE similarity metrics between crest-line points regarding combinations of three and four substrates. In brackets, the minimum achievable MSE value (mm) is shown.

Substrates	NMI	MSE [35]	Substrates	NMI	MSE [35]
BLX SBX DE	−1.291	775.6	BLX SA GM	−1.291	258.3
BLX SBX HS	−1.293	37.0	BLX SA CM	−1.293	259.7
BLX SBX GM	−1.293	295.1	BLX GM CM	−1.290	685.9
BLX SBX SA	**−1.295**	**36.6**	SBX HS DE	−1.292	458.6
BLX SBX CM	−1.290	883.7	SBX HS SA	−1.294	36.7
BLX HS DE	−1.294	36.7	SBX HS GM	−1.280	2952.0
BLX HS GM	−1.293	250.4	SBX HS CM	−1.285	705.6
BLX HS SA	−1.294	254.6	SBX DE SA	−1.292	509.9
BLX HS CM	−1.293	36.8	SBX DE GM	−1.284	2001.5
BLX DE GM	**−1.295**	**36.6**	SBX DE CM	−1.284	2635.9
BLX DE SA	−1.293	261.1	SBX SA GM	−1.290	1100.5
BLX DE CM	−1.292	258.7			
BLX SBX HS DE	**−1.295**	36.6	BLX SA GM CM	−1.292	36.7
BLX SBX HS GM	−1.293	37.4	SBX HS DE SA	−1.292	37.4
BLX SBX HS SA	−1.291	1101.9	SBX HS DE GM	−1.291	37.0
BLX SBX HS CM	−1.293	36.7	SBX HS DE CM	−1.288	296.5
BLX SBX DE GM	−1.295	**36.6**	SBX HS SA GM	−1.291	263.9
BLX SBX DE SA	−1.293	719.4	SBX HS SA CM	−1.289	886.6
BLX SBX DE CM	−1.294	459.8	SBX DE SA GM	−1.292	264.7
BLX SBX GM SA	−1.293	457.1	SBX DE SA CM	−1.288	1106.1
BLX SBX GM CM	−1.289	1355.9	SBX DE GM CM	−1.280	1615.2
BLX SBX SA CM	−1.294	36.6	SBX SA GM CM	−1.286	1120.4
BLX HS DE SA	−1.293	254.0	HS DE SA GM	−1.282	912.2
BLX HS DE GM	−1.293	258.7	HS DE SA CM	−1.281	1520.7
BLX HS DE CM	−1.292	253.0	HS DE GM CM	−1.269	701.5
BLX DE SA GM	−1.292	253.6	HS SA GM CM	−1.272	685.0
BLX DE SA CM	−1.291	256.9	DE SA GM CM	−1.274	3677.0

The best substrate combinations are highlighted in bold.

Finally, we analyze the results of the remaining combinations in Table A.11. The last results confirm the previous remarks, where BXL-α and SBX become essential operators to consider. On the opposite, in those configurations involving one of the two mutation operators, CM yielded to worst results than GM. Similarly, those combinations including SA behave inconsistently. A plausible explanation is the complexity of the calculations required for this operator. As it consumes precious computational resources, the optimization process advances at a slower rate. While most of the configurations are still affected with outlier results due to the stagnation of the algorithm in some runs, three high-level combinations stand out providing the most robust results. Specifically, a five-substrate combination is the one obtaining the lowest averaged MSE. It involves a well-balanced configuration including two crossover operators (BLX-α and SBX), two evolutionary-based operators (HS and DE), and one mutation (GM). This setup will be used as the main design for our proposed CRO-SL for medical IR.

Appendix A.4. Analysis of experimental results

In order to provide an overall analysis of the performance of our proposed CRO-based algorithms against the state-of-the-art evolutionary IR methods, we organized the experimental setup as follows. First, we individually evaluate the performance of each IR approach from an specific optimization problem perspective, in terms of the similarity metric results (NMI in case of intensity-based and MedSE for feature-based approaches). Next, we compare the results of the entire set of algorithms in a standardized framework to evaluate the quality of the registration (using MSE between feature points as the common metric). Note that, as the optimization scheme of the algorithms is non-deterministic, thirty different runs for each IR scenario were performed to avoid execution bias. A statistical analysis was performed over the averaged results for each scenario and complemented with a visual assessment of the results.

Table A.11
Averaged results for the NMI and MSE similarity metrics between crest-line points regarding combinations of five, six, and seven substrates. In brackets, the minimum achievable MSE value (mm) is shown.

Substrates	NMI	MSE [35]	Substrates	NMI	MSE [35]
BLX SBX HS DE GM	**−1.295**	**36.5**	BLX SA HS DE CM	−1.292	37.0
BLX SBX HS DE CM	−1.293	458.9	BLX SA HS GM CM	−1.289	261.8
BLX SBX HS SA GM	−1.291	718.4	BLX SA DE GM CM	−1.291	36.9
BLX SBX HS SA CM	−1.292	680.1	BLX HS DE GM CM	−1.287	37.4
BLX SBX HS GM CM	−1.292	37.3	SBX SA HS DE GM	−1.288	1305
BLX SBX DE SA GM	−1.293	36.6	SBX SA HS DE CM	−1.288	476.1
BLX SBX DE SA CM	−1.291	259.8	SBX SA HS GM CM	−1.288	293.7
BLX SBX DE GM CM	−1.291	458.2	SBX SA DE GM CM	−1.286	1096
BLX SA HS DE GM	−1.292	257.6	SBX HS DE GM CM	−1.288	459.5
BLX SBX HS DE GM CM	−1.293	37.1	BLX SBX SA DE GM CM	**−1.294**	**36.6**
BLX SBX HS DE GM SA	−1.292	473.5	BLX SA HS DE GM CM	−1.288	257.9
BLX SBX HS DE SA CM	−1.292	37.0	SA SBX HS DE GM CM	−1.286	727.8
BLX SBX HS SA GM CM	−1.290	908.2			
BLX SBX HS DE GM SA CM	**−1.293**	**36.7**			

Comparison of intensity-based approaches

Table A.12 shows the mean NMI values for the thirty runs of the intensity-based algorithms, including the averaged ranking for each IR scenario. Considering the results of the similarity metric, every method outperformed the behavior of MS*, which indicates that the complexity of the optimization problem is not trivial and random solutions cannot provide optimal results. r-GA* performed considerably worse than the rest of the algorithms (excluding MS*), ranking in fourth place. SS*, which is currently considered the state of the art in evolutionary IR [13,23], ranked in third place with results substantially better than r-GA*, but was outperformed by both CRO-based approaches proposed in this contribution. CRO* and CRO-SL* delivered a similar performance in terms of ranking: 1.68 and 1.31, respectively. Specifically, CRO* obtained the best results in five occasions, while CRO-SL* outperformed the rest of the algorithms in eleven out of the sixteen scenarios. The difference in terms of NMI between both approaches is minimal in some scenarios but CRO-SL* obtained the lowest rank thanks to its stability during the thirty runs.

Table A.12
Average NMI results for each registration instance.

Scenario	MS*	r-GA*	SS*	CRO*	CRO-SL*
I_1 vs $T_1(I_2)$	−1.1799	−1.2177	−1.2998	−1.3039	**−1.3042**
I_1 vs $T_2(I_2)$	−1.1910	−1.2160	−1.3462	**−1.3608**	−1.3599
I_1 vs $T_3(I_2)$	−1.1773	−1.1926	−1.2794	**−1.3007**	−1.3004
I_1 vs $T_4(I_2)$	−1.1757	−1.1989	−1.2559	−1.2569	**−1.2570**
I_1 vs $T_1(I_3)$	−1.1794	−1.2234	−1.2963	−1.3027	**−1.3030**
I_1 vs $T_2(I_3)$	−1.1915	−1.2219	−1.3406	−1.3595	**−1.3597**
I_1 vs $T_3(I_3)$	−1.1704	−1.1800	−1.2920	−1.2997	**−1.3001**
I_1 vs $T_4(I_3)$	−1.1738	−1.2052	−1.2557	−1.2561	**−1.2562**
I_1 vs $T_1(I_4)$	−1.1836	−1.2043	−1.2863	**−1.2884**	−1.2867
I_1 vs $T_2(I_4)$	−1.1919	−1.2094	−1.3269	−1.3357	**−1.3359**
I_1 vs $T_3(I_4)$	−1.1726	−1.1801	−1.2681	**−1.2869**	−1.2867
I_1 vs $T_4(I_4)$	−1.1801	−1.1995	−1.2492	−1.2496	**−1.2498**
I_2 vs $T_1(I_4)$	−1.1832	−1.2128	−1.2904	−1.2921	**−1.2923**
I_2 vs $T_2(I_4)$	−1.1877	−1.2246	−1.3201	−1.3371	**−1.3371**
I_2 vs $T_3(I_4)$	−1.1744	−1.1893	−1.2798	**−1.2910**	−1.2908
I_2 vs $T_4(I_4)$	−1.1875	−1.2037	−1.2545	−1.2549	**−1.2550**
Average NMI	**−1.1812**	**−1.2050**	**−1.2901**	**−1.2985**	**−1.2984**
Ranking	**5**	**4**	**3**	**1.6875**	**1.3125**

The best substrate combinations are highlighted in bold.

Comparison of feature-based approaches

Accordingly, Table A.13 includes the averaged MedSE values and the corresponding ranking for the feature-based IR methods. In this comparison the differences regarding the similarity metric values are more significant. Recall that MedSE is computed over the crest-lines points instead of the

intensity values, easing the optimization procedure. MSis still an unreliable approach, ranking last with a score of 5.0. The behavior of HE-GA is rather unstable with a poor average performance but decent results in two scenarios. Hence, HE-GA ranks in fourth place (3.63). Overall, SS secured the third rank (3.19) with a better and steadier performance than HE-GA. Once again, the new CRO-based approaches obtained better results than previous outstanding feature-based methods. CRO ranks in second place (1.69) with considerably better results than SS, but CRO-SL stands out as the best feature-based IR approach with a rank of 1.5. Our CRO-SL approach obtained the best results in terms of similarity metric in ten of the sixteen scenarios.

Table A.13
Average MedSE results for each registration instance.

Scenario	MS	HE-GA	SS	CRO	CRO-SL
I_1 vs $T_1(I_2)$	0.966	0.681	0.484	0.452	**0.414**
I_1 vs $T_2(I_2)$	0.958	0.792	0.582	0.480	**0.413**
I_1 vs $T_3(I_2)$	0.963	0.641	0.471	0.414	**0.413**
I_1 vs $T_4(I_2)$	0.963	0.670	0.453	**0.414**	0.452
I_1 vs $T_1(I_3)$	0.975	0.689	0.623	**0.583**	0.596
I_1 vs $T_2(I_3)$	0.967	0.914	0.648	0.598	**0.568**
I_1 vs $T_3(I_3)$	0.971	0.684	0.610	0.568	**0.567**
I_1 vs $T_4(I_3)$	0.971	0.687	0.609	**0.568**	0.596
I_1 vs $T_1(I_4)$	0.978	0.976	0.902	0.899	**0.896**
I_1 vs $T_2(I_4)$	0.971	0.967	0.926	0.903	**0.891**
I_1 vs $T_3(I_4)$	0.974	0.971	0.903	0.892	**0.889**
I_1 vs $T_4(I_4)$	0.975	0.951	0.898	**0.891**	0.896
I_2 vs $T_1(I_4)$	0.971	0.682	0.714	**0.675**	0.685
I_2 vs $T_2(I_4)$	0.964	0.813	0.697	0.685	**0.663**
I_2 vs $T_3(I_4)$	0.965	0.662	0.688	0.666	**0.662**
I_2 vs $T_4(I_4)$	0.965	0.675	0.687	**0.666**	0.684
Average MedSE	**0.969**	**0.742**	**0.681**	**0.647**	**0.643**
Ranking	**5.00**	**3.63**	**3.19**	**1.69**	**1.5**

The best averaged results for each scenario are highlighted in bold.

Comparison of CRO-SL vs previous approaches

Next, we gather all the IR methods for a conjoint comparison using the described common framework. Keep in mind that neither NMI nor MedSE directly correlate with the chosen MSE quality metric, hence the conclusions of this analysis may differ from the previous ones. Thus, Table A.14 reports the MSE results calculated between sets of feature points extracted from the images, and the partial ranking of each IR method for the sixteen scenarios. As mentioned in Section Appendix A.2, the minimum achievable MSE error of each registration scenario accounts for the level of noise and lesion between the compared images. Even if no transformation is applied to the considered images, the distance between crest-line features guiding the error measurement can never be null in any case, as the images are essentially different. Therefore, to clarify the interpretation of the results, the optimal MSE value per scenario is highlighted in brackets in Table A.14. The statistical analysis was performed over the mean MSE results. Ranking, Bonferroni-Dunn's test, and Holm's test results are included in Table A.15.

Table A.14
Detailed results regarding minimum, mean, and standard deviation for the MSE values (in mm) and mean Ranking for each IR instance. Colored rows correspond to the intensity-based implementation of the considered algorithms. Non colored rows correspond to the feature-based implementation. In brackets, the optimal error per scenario is shown.

Algorithm	I_1 vs $T_1(I_2)$ [31]				I_1 vs $T_2(I_2)$ [31]				I_1 vs $T_3(I_2)$ [31]			
	m	μ	sd	R	m	μ	sd	R	m	μ	sd	R
MS*	**31.43**	4640	>99	10	42.18	991	>99	10	53.11	4496	>99	9
MS	36.54	66.26	30	7	39.58	721	>99	9	34.93	446	>99	7
r-GA*	37.32	134	>99	9	39.60	86.43	44	7	43.29	5506	>99	10
HE-GA	42.69	101	47	8	**31.85**	44.27	17	6	**31.96**	42.19	9	6
SS*	36.01	38.14	5.9	6	36.53	38.96	2.9	5	36.78	1755	>99	8
SS	32.06	32.84	2.2	3	32.11	119	>99	8	32.17	41.08	47	3
CRO*	36.35	36.79	0.3	5	36.26	36.78	0.3	4	40.72	41.41	0.5	4
CRO	32.11	32.25	0.1	2	32.10	33.96	5.3	2	32.14	32.27	0.1	2
CRO-SL*	36.01	36.77	0.2	4	36.48	36.77	0.4	3	40.02	41.44	0.4	5
CRO-SL	32.04	**32.21**	0.1	1	32.14	**32.25**	0.1	1	32.11	**32.22**	0.0	1
	I_1 **vs** $T_4(I_2)$ [31]				I_1 **vs** $T_1(I_3)$ [42]				I_1 **vs** $T_2(I_3)$ [42]			
	m	μ	sd	R	m	μ	sd	R	m	μ	sd	R
MS*	34.79	235	>99	9	80.36	4535	>99	10	47.98	149	>99	9
MS	610	1019	>99	10	51.98	485	>99	9	55.09	1079	>99	10
r-GA*	32.19	84.36	91	8	50.61	142	>99	8	45.61	97.75	81	8
HE-GA	32.88	59.23	28	7	61.65	122	48	7	**42.91**	61.01	24	6
SS*	32.65	32.96	0.7	6	50.20	61.34	50	6	43.65	46.36	3.9	5
SS	**32.00**	32.25	0.1	3	**42.96**	58.69	59	5	43.08	83.82	>99	7
CRO*	32.67	32.79	0.1	4	51.23	51.77	0.3	4	43.54	43.73	0.1	3
CRO	32.14	32.24	0.0	2	43.01	**43.29**	0.2	1	42.94	44.06	2.9	4
CRO-SL*	32.75	32.79	0.0	5	51.60	51.75	0.1	3	43.59	43.72	0.1	2
CRO-SL	32.18	**32.24**	0.1	1	43.01	43.37	0.2	2	42.96	**43.37**	0.2	1

(*continued on next page*)

Table A.14 *(continued)*

	I_1 **vs** $T_3(I_3)$ [42]				I_1 **vs** $T_4(I_3)$ [42]				I_1 **vs** $T_1(I_4)$ [46]			
	m	μ	sd	R	m	μ	sd	R	m	μ	sd	R
MS*	72.98	8896	>99	10	53.92	3809.4	>99	10	54.09	2062	>99	10
MS	51.00	376	>99	8	620	859	>99	9	58.09	509	>99	9
r-GA*	66.00	8470	>99	9	42.93	86.11	53	7	51.06	151	>99	8
HE-GA	56.52	74.44	18	7	58.74	114	42	8	60.07	119	51	7
SS*	55.22	57.85	3.2	6	44.68	45.18	0.3	6	52.53	53.67	1.3	5
SS	43.06	50.95	41	3	42.95	43.44	0.6	3	46.79	61.77	77	6
CRO*	55.78	56.53	0.4	5	44.61	45.17	0.3	5	52.78	53.24	0.3	4
CRO	**42.97**	43.45	0.2	2	**42.87**	43.39	0.3	2	46.66	48.85	6.0	2
CRO-SL*	56.25	56.52	0.1	4	45.03	45.15	0.1	4	50.54	52.97	0.7	3
CRO-SL	43.00	**43.34**	0.2	1	43.12	**43.31**	0.2	1	**46.43**	**47.19**	0.4	1
	I_1 **vs** $T_2(I_4)$ [46]				I_1 **vs** $T_3(I_4)$ [46]				I_1 **vs** $T_4(I_4)$ [46]			
	m	μ	sd	R	m	μ	sd	R	m	μ	sd	R
MS*	48.51	3020	>99	10	61.52	8681.1	>99	9	51.39	183	>99	9
MS	826	1096	>99	9	52.03	376	>99	8	599	884	>99	10
r-GA*	58.92	124	98	6	60.76	9428.7	>99	10	**46.44**	82.74	45	7
HE-GA	843	1011	>99	8	51.18	78.92	25	7	51.35	92.56	30	8
SS*	46.48	48.12	1.5	4	54.32	60.10	5.2	5	46.65	47.58	0.5	4
SS	46.52	154	>99	7	46.87	64.26	68	6	46.67	**47.33**	0.5	1
CRO*	**46.02**	46.37	0.1	2	57.46	58.02	0.3	4	47.03	47.70	0.3	6
CRO	46.77	48.80	3.4	5	46.80	47.57	0.6	2	46.64	47.37	0.6	3
CRO-SL*	46.21	**46.36**	0.1	1	53.96	57.83	0.8	3	47.55	47.68	0.0	5
CRO-SL	46.59	47.42	0.6	3	**46.56**	**47.14**	0.4	1	46.74	47.37	0.4	2

Algorithm	I_2 vs $T_1(I_4)$ [28]				I_2 vs $T_2(I_4)$ [28]			
	m	μ	sd	R	m	μ	sd	R
MS*	41.98	4693	>99	10	31.77	2183	>99	10
MS	31.71	460	>99	8	32.89	681	>99	9
r-GA*	32.15	3293	>99	9	32.20	68.26	33	8
HE-GA	33.59	94.40	35	7	29.94	46.64	41	7
SS*	34.80	35.66	0.5	5	30.36	33.07	3.7	5
SS	28.54	44.21	79	6	28.61	35.27	35	6
CRO*	35.06	35.54	0.2	3	30.25	30.47	0.1	4
CRO	**28.32**	28.79	0.3	2	28.61	29.01	0.6	2
CRO-SL*	35.54	35.62	0.0	4	30.24	30.44	0.2	3
CRO-SL	28.53	**28.74**	0.1	1	**28.53**	**28.75**	0.1	1
	I_2 **vs** $T_3(I_4)$ [28]				I_2 **vs** $T_4(I_4)$ [28]			
	m	μ	sd	R	m	μ	sd	R
MS*	34.22	10045	>99	10	30.34	88.60	56	9
MS	35.28	148	>99	8	574	676	>99	10
r-GA*	33.38	3656	>99	9	29.53	57.89	32	7
HE-GA	29.88	37.67	14	4	31.34	78.25	33	8
SS*	38.85	43.70	9.6	7	**28.44**	29.69	0.5	6
SS	**28.46**	28.80	0.1	2	28.57	28.75	0.1	2
CRO*	39.11	40.66	0.5	5	29.36	29.64	0.1	5
CRO	28.57	28.82	0.1	3	28.59	28.76	0.1	3
CRO-SL*	38.36	40.79	0.5	6	29.57	29.61	0.0	4
CRO-SL	28.55	**28.72**	0.1	1	28.56	**28.67**	0.1	1

The best averaged results for each scenario are highlighted in bold.

In general, feature-based approaches outperform their intensity-based counterparts in most of the occasions. The main reason for this behavior is that the optimization design of feature-based approaches allows them to achieve a better accuracy in less time. Hence, feature-based approaches usually obtain the best minimum results, at the expense of increasing the amount of variance. On the other hand, the results provided by most of the intensity-based methods (but MS* and r-GA*) are more robust, showing lower standard deviation values than feature-based approaches. Fig. A.7 provides an overview of the performance and the variability of the different methods regarding the sixteen scenarios.

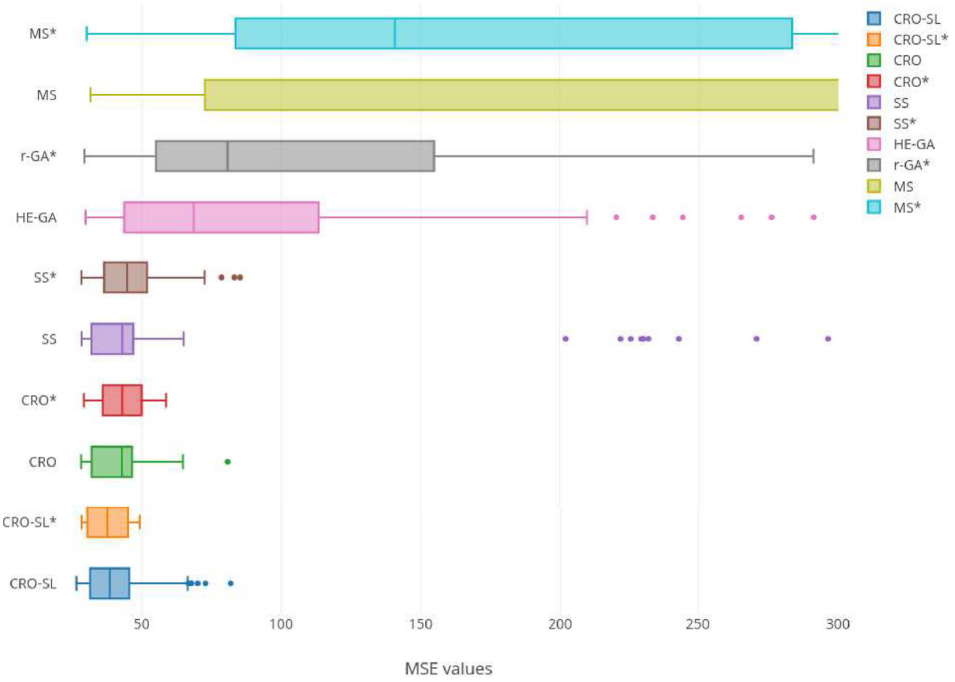

Fig. A.7. Box plot of the MSE results obtained by each IR method. Results with an MSE value greater than 300 were omitted to increase the visualization quality.

Multi-start approaches fall to the lowest positions in the ranking. In particular, MS* and MS obtained the worst ranks (9.63 and 8.75, respectively) due to scoring the largest MSE values consistently. They presented a high variance, obtaining mean MSE values at least one order of magnitude away from the best solutions in most of the scenarios. Similarly, both genetic approaches performed with an unstable behavior but were able to deliver good minimum results. r-GA* outperformed both MS* and MS, and ranked in the eighth place (8,13). Despite its mediocre mean performance, HE-GA was able to achieve the best overall minimum result in three of the sixteen scenarios considering all the methods, with a ranking score of 6.94.

SS* and SS performed considerably better than the previous methods, behaving consistently except in four different scenarios due to outlier results. Thus, SS-based methods reached medium ranking scores (5.56 and 4.44, respectively). Even though both methods perform similarly in terms of averaged MSE and failed to converge in some occasions, SSranked above SS* by delivering better minimum outcomes despite a high standard deviation.

Regarding the basic CRO approaches, CRO* was able to obtain better mean MSE results that both SS approaches, but when considering minimum MSE values, it is not able to reach their accuracy level. On the other hand, the feature-based approach CRO outperformed its intensity-based counterpart and the current state-of-the art evolutionary IR method in this comparison (SS). In terms of mean MSE values, CRO was able to outperform SS in twelve scenarios. Therefore, CRO* ranked in fourth place with 4.19 while CRO was second with a rank of 2.44.

The substrate-layer version of CRO performed slightly better than the basic design, so the conclusions are similar. CRO-SL* was the intensity-based approach with lowest standard deviation values, improving the results obtained by CRO* regarding mean MSE in most of the scenarios. This method ranked in third place with 3.69, being the best intensity-based method. The feature-based approach CRO-SL had excellent results according to the mean MSE values, reaching the best result in thirteen instances and obtaining the first place in the rank with 1.25. We can thus conclude that CRO-SL is the best IR method in the comparison for both intensity- and feature-based approaches. The good performance of our proposal can be explained by the trade-off obtained by the combination of different exploration operators.

Summarizing, all the CRO-based proposals were able to outperform the current state-of-the-art evolutionary IR methods with the feature-based approaches obtaining the best performance. The basic version CRO was able to deliver accurate results at the expense of higher variability, while the substrate-layer design CRO-SL excelled thanks to its stability and robustness. This good performance of the CRO-SL in the problem is explained by different factors. Among them, we highlight the fact that the CRO-based approaches introduce a global search pattern, instead of the local approximation of the classical methods. Moreover, the CRO-SL provides an excellent balance of intensification-diversification in the search, together with the inclusion of new search patterns in a single population. This gives the algorithm a robust search characteristic, which mainly depend on the different substrates included in the simulated reef. In addition, the CRO-SL has shown a fast convergence to a high quality solution, responding very well to the quality-computation time defined in the experimental setting, independently of the landscape of the problem.

According to the statistical analysis, the result of applying Friedman's test is $\chi_F^2 = 122.38$, and the corresponding p-value is $< 10^{-22}$. Given that the p-value is lower than the considered level of significance ($\alpha = 0.01$), the test concludes that there are significant differences among the results.

Table A.15
Ranking obtained through Friedman's test and statistical p-values with CRO-SL* as control method for Bonferroni's test, and Holm's test according to the mean MSE value.

	Mean Rank	Bonferroni-Dunn p	Holm p
CRO-SL	1.25	–	–
CRO	2.44	1	0.27
CRO-SL*	3.69	0.21	< 0.05
CRO*	4.19	0.05	< 0.05
SS	4.44	0.03	< 0.01
SS*	5.56	$< 10^{-04}$	$< 10^{-04}$
HE-GA	6.94	$< 10^{-07}$	$< 10^{-07}$
r-GA*	8.13	$< 10^{-09}$	$< 10^{-10}$
MS	8.75	$< 10^{-11}$	$< 10^{-11}$
MS*	9.63	$< 10^{-14}$	$< 10^{-14}$

We complemented the analysis with Holm's test (Table A.15) comparing CRO-SL (control) with the rest of the methods. The p-value results for both tests reveal that CRO-SL present significant differences with SS-based, genetic algorithm-based, and MS-based approaches.

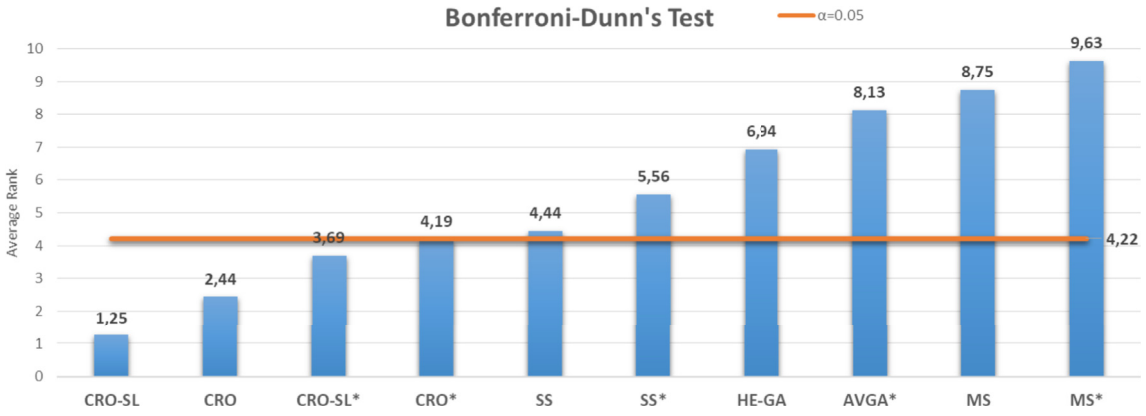

Fig. A.8. Bonferroni-Dunn's graphic corresponding to the test results.

Finally, we include a comparison of different registration solutions obtained by the IR methods in one of the most complex scenarios to provide a visual assessment of the quality of the results. Fig. A.9 presents a visualization of the overlapping between the images regarding the best solution obtained by each algorithm. Considering the worst performing methods, the intensity-based version of MS achieved a better overlapping than its feature-based counterpart, even though it obtained a higher MSE. A similar conclusion can be obtained for the behavior of the genetic approaches, as the resulting transformations present a slight misalignment. Regarding the best performing methods, SS-based and CRO-based methods achieved a good overlapping between images. [3] It is appreciable how CRO and CRO-SL (with similar MSE) improve the accuracy of the alignment provided by SS.

[3] If an optimal transformation is achieved, the visualization shows a brain of interlaced colors, where yellow (light gray) represents the transformed image and blue (dark gray) depicts the reference image.

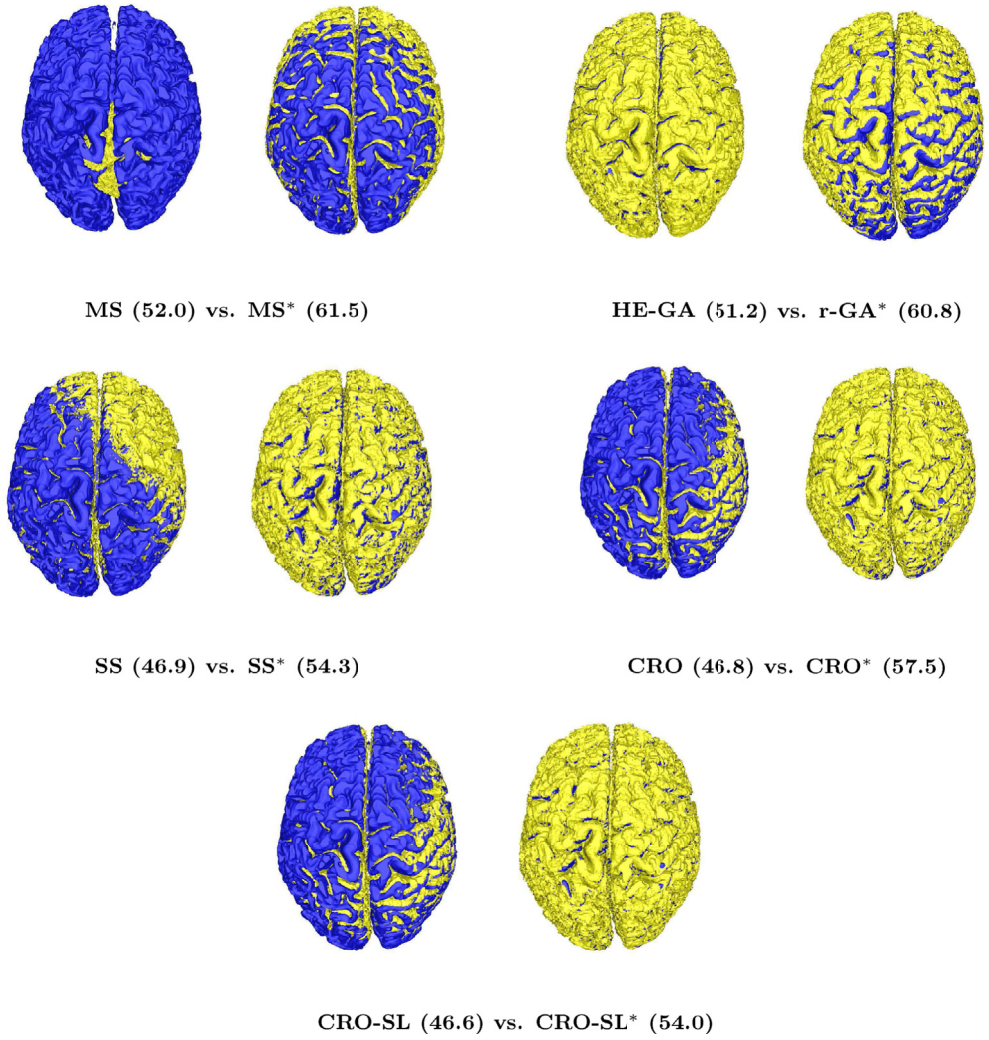

Fig. A.9. Visual results of the overlapping between model (blue) and the scene (yellow). Figures provide a comparison of the results obtained by the feature-based (left) and the intensity-based (right) methods regarding the 11*th* scenario (I_1 vs. $T_3(I_4)$). The solutions correspond to the minimum MSE value obtained by the corresponding algorithm, which is included in parenthesis. The reader is referred to the web version of the article for interpretation of the color references in this figure.

Appendix B. Rigid multi-modal experimental study

We present in this appendix a brief extension of the results achieved by the compared IR methods in the multi-modal study. As the problem instances extracted from the RIRE dataset relate different image modalities of the same patient, affine transformations introduce undesired deformations during the registration procedure. This kind of transformation model was considered to assess the optimization behavior of the methods in a more challenging problem. Therefore, this appendix complements our contribution by presenting the results of applying a rigid body transformation, which corresponds to an anatomically correct transformation for the considered registration problem.

Table B.16 present the results of this simplified study. Here we also include the results available in the RIRE system for another outstanding global optimization approach, CMTK.[4] Note that the results of CMTK and the rest of the methods are not directly comparable in terms of optimization performance, as the execution conditions of the registration frameworks differ. Nevertheless, the results obtained by this state-of-the-art method have been included as a reference for its informative value and outstanding accuracy.

[4] It was not possible to include another relevant method (Zikic et al.'s MRF [7]) in the current study. The results provided in that contribution are not directly comparable as the considered instances differ among the studies.

Table B.16
Detailed results for the registration error (mm) of each IR scenario in the RIRE dataset considering a rigid body transformation. Evolutionary methods include the averaged result of the 30 runs and the standard deviation. The last row shows the number of times an algorithm obtains the best outcome.

Scenario	Simplex	FLIRT	CMTK	CRO-SL*		CRO*		SS*		ASGD	
				μ	sd	μ	sd	μ	sd	μ	sd
PD-P101	9.02	3.19	2.94	**2.88**	0.2	3.18	0.2	3.08	0.2	3.29	0.3
PD-P102	9.20	2.38	2.32	**2.29**	0.1	2.51	0.2	2.54	0.2	2.32	0.1
PD-P103	4.10	3.35	**3.11**	3.21	0.1	3.31	0.3	3.32	0.1	3.23	0.1
PD-P104	4.43	**2.02**	2.02	2.20	0.1	2.24	0.2	2.42	0.2	2.21	0.1
T1-P101	24.29	1.96	1.93	**1.79**	0.1	2.07	0.2	1.96	0.2	4.67	1.2
T1-P102	9.05	**1.51**	1.54	1.56	0.1	1.76	0.2	1.76	0.2	1.54	0.1
T1-P103	2.68	2.48	**2.37**	2.46	0.1	2.48	0.2	2.44	0.2	2.37	0.1
T1-P104	3.72	1.43	1.38	1.54	0.2	**1.37**	0.1	1.49	0.2	1.51	0.1
T1-P105	13.41	2.64	2.36	**2.34**	0.2	2.54	0.4	2.48	0.3	2.48	0.1
T1-P106	8.64	1.86	1.98	**1.84**	0.2	1.98	0.1	1.99	0.2	2.75	0.6
T1-P107	21.51	1.20	**0.97**	1.11	0.1	1.32	0.2	1.38	0.2	1.35	0.3
T1-P108	11.23	1.82	1.67	**1.56**	0.1	1.82	0.2	1.68	0.2	1.85	0.5
T1-P109	3.87	**1.36**	1.46	1.52	0.1	1.71	0.2	1.77	0.1	1.60	0.1
T2-P101	12.25	2.99	2.70	**2.69**	0.1	2.81	0.2	2.97	0.2	2.78	0.2
T2-P102	8.88	2.04	1.82	1.98	0.1	1.90	0.2	1.99	0.2	**1.79**	0.1
T2-P104	4.34	1.78	1.77	1.74	0.1	1.84	0.2	1.69	0.2	**1.63**	0.1
T2-P105	14.06	2.88	2.75	2.84	0.2	**2.74**	0.1	2.78	0.2	2.97	0.4
T2-P106	8.18	2.71	2.63	2.47	0.1	2.53	0.1	2.50	0.2	**2.36**	0.1
T2-P107	6.50	**1.98**	1.98	2.20	0.1	2.28	0.2	2.20	0.2	2.32	0.1
T2-P108	11.00	2.39	2.04	**1.60**	0.1	1.80	0.2	1.92	0.2	1.80	0.2
T2-P109	4.34	2.09	1.99	**1.67**	0.3	1.77	0.2	1.83	0.1	1.86	0.1
# wins	–	4	3	9		2		–		3	

The best averaged results for each scenario are highlighted in bold.

The analysis of the rigid results provide a similar insight to the affine ones. Simplex delivered the worst performance among all the methods. Meanwhile, ASGD, CRO*, and SS* obtained a standard and robust performance. FLIRT and CMTK provided slightly more accurate results than the previous methods. The achievement of the best minimum error is distributed among these methods, excepting Simplex and SS*. In contrast, CRO-SL stands out in the comparison by reaching the best accuracy in 9 of the 21 instances.

Nevertheless, the simplicity of the search space denote many similarities among the compared methods. Therefore, the results of the statistical tests shown in Table B.17 verify the lack of significant differences between CRO-SL*, CMTK, FLIRT, CRO*, and ASGD at a confidence level of $\alpha = 0.05$.

Table B.17
Ranking obtained through Friedman's test and statistical p-values with CRO-SL* as control method for Bonferroni's test, and Holm's test according to the averaged registration error.

	Rank	Bonferroni p	Holm p
CRO-SL*	2.38	–	–
CMTK	2.48	1	0.89
ASGD	3.86	0.16	0.05
CRO*	4.00	0.09	0.05
FLIRT	4.05	0.07	0.05
SS*	4.24	0.03	0.03
Simplex	7.00	$< 10^{-11}$	$< 10^{-11}$

In a direct comparison between CRO-SL and CMTK, our proposal stands out by a close margin, obtaining the best results in 11 of the 21 instances. As a conclusion, both experimental studies performed on the RIRE dataset demonstrate the suitability of our proposal as a robust and efficient optimization algorithm to deal with real-world applications.

References

[1] B. Zitová, J. Flusser, Image registration methods: a survey, Image Vis Comput. 21 (11) (2003) 977–1000, https://doi.org/10.1016/S0262-8856(03)00137-9.

[2] A.A. Goshtasby, 2-D and 3-D Image Registration, Wiley Interscience, 2005, https://doi.org/10.1002/0471724270.

[3] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256, https://doi.org/10.1109/34.121791.

[4] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, second ed., Cambridge University Press, 1995.

[5] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, Neuroimage 17 (2) (2002) 825–841, https://doi.org/10.1016/S1053-8119(02)91132-8.

[6] S. Klein, J. Pluim, M. Staring, M. Viergever, Adaptive stochastic gradient descent optimisation for image registration, Int. J. Comput. Vis. 81 (2009) 227–239.

[7] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Kamen, N. Pargios, N. Navab, Linear intensity-based image registration by Markov random fields and discrete optimization, Med. Image Anal. 14 (2010) 550–562.

[8] T. Ting, X.-S. Yang, S. Cheng, K. Huang, Hybrid Metaheuristic Algorithms: Past, Present, and Future, vol. 585, Springer International Publishing, 2015, https://doi.org/10.1007/978-3-319-13826-8_4.

[9] R. He, P.A. Narayana, Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images, Comput. Med. Imag. Graph. 26 (4) (2002) 277–292.

[10] R. Panda, S. Agrawal, M. Sahoo, R. Nayak, A novel evolutionary rigid body docking algorithm for medical image registration, Swarm Evolut. Comput. 33 (2018) 108–118, https://doi.org/10.1016/j.swevo.2016.11.002.

[11] F. Glover, M. Laguna, R. Martí, Scatter search, in: A. Ghosh, S. Tsutsui (Eds.), Advances in Evolutionary Computation: Theory and Applications, Springer-Verlag, New York, 2003, pp. 519–537.

[12] S. Damas, O. Cordon, J. Santamaria, Medical image registration using evolutionary computation: an experimental survey, IEEE Comput. Intell. Mag. 6 (4) (2011) 26–42, https://doi.org/10.1109/Mci.2011.942582.

[13] A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem, Intensity-based image registration using scatter search, Artif. Intell. Med. 60 (3) (2014) 151–163, https://doi.org/10.1016/j.artmed.2014.01.006.

[14] S. Salcedo-Sanz, J. Del Ser, I. Landa-Torres, S. Gil-López, J.A. Portilla-Figueras, The coral reefs optimization algorithm: a novel metaheuristic for efficiently solving optimization problems, Sci. World J. 2014 (2014) 1–15.

[15] S. Salcedo-Sanz, J. Muñoz-Bulnes, M. Vermeij, New coral reefs-based approaches for the model type selection problem: a novel method to predict a Nation's future energy demand, Int. J. Bio-Inspired Comput. 10 (3) (2017), https://doi.org/10.1504/IJBIC.2017.086698.

[16] D.L. Collins, A.P. Zijdenbos, V. Kollkian, J.G. Sled, N.J. Kabani, C.J. Holmes, A.C. Evans, Design and construction of a realistic digital brain phantom, IEEE Trans. Med. Imag. 17 (1998) 463–468, https://doi.org/10.1109/42.712135.

[17] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer Jr., R. Kessler, R. Maciunas, C. Barillot, D. Lemoine, A. Collignon, et al., Comparison and evaluation of retrospective intermodality brain image registration techniques, J. Comput. Assist. Tomogr. 21 (4) (1997) 554.

[18] M. Andreetto, G.M. Cortelazzo, L. Lucchese, Frequency domain registration of computer tomography data, in: Proceedings - 2nd International Symposium on 3D Data Processing, Visualization, and Transmission. 3DPVT 2004, 2004, pp. 550–557, https://doi.org/10.1109/TDPVT.2004.1335287.

[19] O. Zvitia, A. Mayer, R. Shadmi, S. Miron, H.K. Greenspan, Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling, IEEE Trans. Med. Imag. 29 (1) (2010) 132–145, https://doi.org/10.1109/TMI.2009.2029097.

[20] M. Svedlow, C. McGillem, P. Anuta, Experimental examination of similarity measures and preprocessing methods used for image registration, (A), Indiana, USA, in: P.H. Swain, D.B. Morrison, D.E. Parks (Eds.), Symposium on Machine Processing of Remotely Sensed Data, vol. 4, Purdue University, West Lafayette, Ind, 1976, 4A9–4A17.

[21] M.A. Audette, F.P. Ferrie, T.M. Peters, An algorithmic overview of surface registration techniques for medical imaging, Med. Image Anal. 4 (3) (2000) 201–217, https://doi.org/10.1016/S1361-8415(00)00014-1.

[22] A. Valsecchi, S. Damas, J. Santamaria, Evolutionary intensity-based medical image registration: a review, Curr. Med. Imag. Rev. 9 (4) (2013) 283–297.

[23] E. Bermejo, A. Valsecchi, S. Damas, O. Cordon, Bacterial Foraging Optimization for intensity-based medical image registration, in: IEEE Congress on Evolutionary Computation, CEC 2015-Proceedings, IEEE, 2015, pp. 2436–2443, https://doi.org/10.1109/CEC.2015.7257187.

[24] J.A. Nelder, R.A. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 313–380.

[25] S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R.K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, P.M. Matthews, Advances in functional and structural MR image analysis and implementation as FSL, Neuroimage 23 (2004) S208–S219, https://doi.org/10.1016/j.neuroimage.2004.07.051.

[26] T. Rohlfing, Multimodale Datenfusion für die bildgesteuerte Neurochirurgie und Strahlentherapie (In German), Ph.D. thesis, Technische Universität Berlin, 2000.

[27] C. Studholme, D.L.G. Hill, D.J. Hawkes, Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures, Med. Phys. 24 (1) (1997) 25, https://doi.org/10.1118/1.598130.

[28] A. Valsecchi, S. Damas, J. Santamaria, L. Marrakchi-Kacem, Genetic algorithms for Voxel-based medical image registration, in: 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), 2013, pp. 22–29, https://doi.org/10.1109/CIMI.2013.6583853.

[29] M. Takahashi, H. Kita, A crossover operator using independent component analysis for real-coded genetic algorithms, in: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), vol. 1, 2001, pp. 643–649, https://doi.org/10.1109/CEC.2001.934452.

[30] F. Glover, Heuristic for integer programming using surrogate constraints, Decis. Sci. J. 8 (1977) 156–166.

[31] J. Santamaría, O. Cordón, S. Damas, I. Aleman, M. Botella, A scatter search-based technique for pair-wise 3D range image registration in forensic anthropology, Soft Comput. 11 (9) (2007) 819–828, https://doi.org/10.1007/s00500-006-0132-0.

[32] M. Lozano, F. Herrera, N. Krasnogor, D. Molina, Real-coded memetic algorithms with crossover hill-climbing, Evol. Comput. 12 (3) (2004) 273–302, https://doi.org/10.1162/1063656041774983.

[33] J. Santamaría, O. Cordón, S. Damas, J.M. García-Torres, A. Quirin, Performance evaluation of memetic approaches in 3D reconstruction of forensic objects, Soft Comput. 13 (8–9) (2009) 883–904, https://doi.org/10.1007/s00500-008-0351-7.

[34] S.M. Yamany, M.N. Ahmed, A. a. Farag, A new genetic-based technique for matching 3-D curves and surfaces, Pattern Recogn. 32 (10) (1999) 1817–1820, https://doi.org/10.1016/S0031-3203(99)00060-6.

[35] H.G. Beyer, K. Deb, On self-adaptive features in real-parameter evolutionary algorithms, IEEE Trans. Evol. Comput. 5 (3) (2001) 250–270, https://doi.org/10.1109/4235.930314.

[36] Salcedo-Sanz, A review on the coral reefs optimization algorithm: new development lines and current applications, Progr. Artif. Intel. 6 (1) (2017) 1–15, https://doi.org/10.1007/s13748-016-0104-2.

[37] M.J.A. Vermeij, Substrate composition and adult distribution determine recruitment patterns in a Caribbean brooding coral, Mar. Ecol. Prog. Ser. 295 (2005) 123–133, https://doi.org/10.3354/meps295123.

[38] S. Salcedo-Sanz, C. Camacho-Gomez, D. Molina, F. Herrera, A coral reefs optimization algorithm with substrate layers and local search for large scale global optimization, in: 2016 IEEE Congress on Evolutionary Computation (CEC), 2016, pp. 3574–3581, https://doi.org/10.1109/CEC.2016.7744242.

[39] S. Salcedo-Sanz, C. Camacho-Gómez, R. Mallol-Poyato, S. Jiménez-Fernández, J. Del Ser, A novel Coral Reefs Optimization algorithm with substrate layers for optimal battery scheduling optimization in micro-grids, Soft Comput. 20 (11) (2016) 4287–4300, https://doi.org/10.1007/s00500-016-2295-7.

[40] S. Salcedo-Sanz, C. Camacho-Gómez, A. Magdaleno, E. Pereira, A. Lorenzana, Structures vibration control via tuned mass dampers using a co-evolution Coral Reefs optimization algorithm, J. Sound Vib. 393 (2017) 62–65, https://doi.org/10.1016/j.jsv.2017.01.019.

[41] S. Salcedo-Sanz, Modern meta-heuristics based on nonlinear physics processes: a review of models and design procedures, Phys. Rep. 655 (2016) 1–70, https://doi.org/10.1016/j.physrep.2016.08.001.

[42] C. Studholme, D.L.G. Hill, D.J. Hawkse, An overlap invariant entropy measure of 3D medical image alignment, Pattern Recogn. 32 (1999) 71–86.

[43] E. Bermejo, M. Chica, S. Salcedo-Sanz, O. Cordon, Coral reef optimization for intensity-based Medical Image registration, in: IEEE Congress on Evolutionary Computation, CEC 2017-Proceedings, IEEE, 2017, pp. 533–540, https://doi.org/10.1109/CEC.2017.7969357.

[44] R. Storn, K. Price, Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359, https://doi.org/10.1023/A:1008202821328.

[45] M.B. Trabia, X.B. Lu, A fuzzy adaptive simplex search optimization algorithm, J. Mech. Des. 123 (2) (2001) 216, https://doi.org/10.1115/1.1347991.

[46] K. Deb, R.B. Agrawal, Simulated binary crossover for continuous search space, Complex Syst. 9 (1994) 1–34, doi:10.1.1.26.8485Cached.

[47] H.-P. Schwefel, Collective phenomena in evolutionary systems, in: P. Checkland, I. Kiss (Eds.), Problems of Constancy and Change - the Complementarity of Systems Approaches to Complexity, Proc. 31st Annual Meeting, vol. 2, International Society for General System Research, Budapest, 1987, pp. 1025–1033.

[48] X. Yao, Y. Liu, G. Lin, Evolutionary programming made faster, IEEE Trans. Evol. Comput. 3 (2) (1999) 82–102, https://doi.org/10.1109/4235.771163.

[49] J.C. Sprott, Strange attractors: creating patterns in Chaos, Am. J. Phys. 63 (5) (1995) 477, https://doi.org/10.1119/1.17885.

[50] P. Grassberger, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (5) (1983) 346–349, https://doi.org/10.1103/PhysRevLett.50.346.

[51] S. Garcia, F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, J. Mach. Learn. Res. 9 (2008) 2677–2694.

[52] M. Friedman, A Comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92, https://doi.org/10.1214/aoms/1177731944.

[53] O.J. Dunn, Multiple comparisons among means, J. Am. Stat. Assoc. 56 (293) (1961) 52–64, https://doi.org/10.2307/2282330.

[54] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (2) (1979) 65–70, https://doi.org/10.2307/4615733.

[55] P. Rogelj, S. Kovacic, J. Gee, Validation of a nonrigid registration algorithm for multimodal data, in: M. Sonka, J.M. Fitzpatrick (Eds.), Proc. SPIE 4684, Medical Imaging, Image Processing, San Diego, USA, 2002, pp. 299–307, https://doi.org/10.1117/12.467170.

[56] O. Monga, S. Benayoun, O. Faugeras, From partial derivatives of 3-D density images to ridge lines, in: Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1808, IEEE, Champaign, Illinois, USA, 1992, pp. 354–359, https://doi.org/10.1109/CVPR.1992.223165.

[57] T. Rohlfing, Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable, IEEE Trans. Med. Imag. 31 (2) (2012) 153–163, https://doi.org/10.1109/TMI.2011.2163944.

[58] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P.W. Pluim, Elastix: a toolbox for intensity-based medical image registration, IEEE Trans. Med. Imag. 29 (1) (2010) 196–205, https://doi.org/10.1109/TMI.2009.2035616.

[59] O. Cordón, S. Damas, J. Santamaría, Feature-based image registration by means of the chc evolutionary algorithm, Image Vis Comput. 24 (5) (2006) 525–533, https://doi.org/10.1016/j.imavis.2006.02.002.

[60] J. Santamaría, O. Cordón, S. Damas, A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling, Comput. Vis. Image Understand. 115 (9) (2011) 1340–1354, https://doi.org/10.1016/j.cviu.2011.05.006.