Evolutionary Diversity Optimization Using Multi-Objective Indicators

Aneta Neumann
The University of Adelaide
School of Computer Science
Adelaide, Australia

Wanru Gao
The University of Adelaide
School of Computer Science
Adelaide, Australia

Markus Wagner
The University of Adelaide
School of Computer Science
Adelaide, Australia

Frank Neumann
The University of Adelaide
School of Computer Science
Adelaide, Australia

ABSTRACT

Evolutionary diversity optimization aims to compute a set of solutions that are diverse in the search space or instance feature space, and where all solutions meet a given quality criterion. With this paper, we bridge the areas of evolutionary diversity optimization and evolutionary multi-objective optimization. We show how popular indicators frequently used in the area of multi-objective optimization can be used for evolutionary diversity optimization. Our experimental investigations for evolving diverse sets of TSP instances and images according to various features show that two of the most prominent multi-objective indicators, namely the hypervolume indicator and the inverted generational distance, provide excellent results in terms of visualization and various diversity indicators.

CCS CONCEPTS

• **Theory of computation** → **Evolutionary algorithms**;

KEYWORDS

Diversity, evolutionary algorithms, features

1 INTRODUCTION

Evolutionary algorithms have been used for a wide range of optimization problems and to discover novel designs for various engineering problems [6, 9]. Diversity plays a crucial role when designing evolutionary algorithms as it often prevents the algorithms from premature convergence. In recent years, evolutionary diversity optimization has gained increasing attention [1, 13, 23, 30, 31]. Evolutionary diversity optimization uses an evolutionary algorithm in order to compute a diverse set of solutions that all fulfill given

quality criteria. Presenting decision makers with such alternative designs that are all of good quality gives them a variety of design choices and helps to better understand the space of good solutions for the problem at hand. Related to evolutionary diversity optimization is the concept of novelty search [27, 28]. Here evolutionary algorithms are used to discover new designs without focusing on an objective. The goal of novelty search is to explore designs that are different to the ones previously obtained. This paper focuses on evolutionary diversity optimization. We are interested in computing a diverse set of high quality solutions that can be presented to a decision maker.

Arguably, the most prominent area of evolutionary computation where a diverse set of solutions is sought is evolutionary multi-objective optimization [8]. Given a set of usually conflicting objective functions, the goal is to compute a set of solutions representing the different trade-offs of the considered functions. Evolutionary algorithms have been widely applied to multi-objective optimization problems and it is one of the key success areas for applying evolutionary algorithms. Over the years, many evolutionary multi-objective algorithms have been developed. Popular algorithms, among many others, are NSGA-II [10], NSGA-III[15], MOEA/D [33], and IBEA [34]. Making them applicable to the area of evolutionary diversity optimization provides huge potential for high performing evolutionary diversity optimization approaches. With this paper, we bridge the areas of evolutionary diversity optimization and evolutionary multi-objective optimization. We consider popular indicators from the area of evolutionary multi-objective optimization and show how to make them applicable in the area of evolutionary diversity optimization.

Ulrich and Thiele [31] have introduced the framework for evolutionary diversity optimization. They studied how to evolve diverse sets of instances for single-objective problems to the underlying search space. Furthermore, this diversity optimization approach has been introduced into multi-objective search [30]. In [13], an evolutionary diversity optimization process has been introduced to evolve instances of the Traveling Salesperson problem (TSP) based on given problem features. This approach evolves TSP instances that are hard or easy to solve for a given algorithm, and diversity is measured according to a weighted distribution in terms of the differences in feature values. Afterwards, the approach has been adapted in order to create variations of a given image that are close to it but differ in terms of the chosen image features [1].

An important question that arises when using evolutionary diversity optimization for more than one criterion or feature is how to measure the diversity of a given set of solutions. The weighted contribution approach used in [1, 12] has the disadvantage that it

heavily depends on the chosen weightening of the features and does not distribute that well for two or three dimensions. In [23], an evolutionary diversity optimization approach has been introduced that aims to minimize the discrepancy of the solution set in the feature space. It has been shown that using the star discrepancy as a diversity measure achieves sets of higher diversity than the previous approaches using weighted contributions.

In this paper, we show how to use popular indicators from the area of evolutionary multi-objective optimization for evolutionary diversity optimization. Indicators play a prominent role in the area of evolutionary multi-objective optimization and are frequently used to assess the quality of solution sets produced by evolutionary multi-objective algorithms [34, 35]. Based on the evaluation of this indicator the selection for survival is carried out. We show how to adapt popular indicators in the area of evolutionary multi-objective optimization to evolutionary diversity optimization. We study important indicators such as the hypervolume indicator (HYP), the inverted generational distance (IGD), and the additive epsilon approximation (EPS), and compare them in terms of their ability to lead to high quality and diverse sets of solutions.

We investigate these indicators for the problems of evolving TSP instances and constructing diverse sets of images as already studied in the literature. Our results show that HYP and IGD are well suited for evolutionary diversity optimization. They obtain the best results for their respective indicator and also obtain sets of solutions of a better discrepancy when comparing them to the discrepancy-based approach given in [23].

The outline of the paper is as follows. In Section 2, we introduce evolutionary diversity optimization and present our approach of using multi-objective indicators in this context. In Sections 3 and 4, we describe our diversity optimization for two problems: diverse sets of images and diverse sets of TSP instances. Finally, we draw some conclusions.

2 INDICATOR-BASED DIVERSITY OPTIMIZATION

Let $I \in X$ be a search point in a given search space X, $f\colon X \to \mathbb{R}^d$ a function that assigns to each search point a feature vector and $q\colon X \to \mathbb{R}$ be a function assigning a quality score to each $x \in X$ [3]. Diversity is defined in terms of a function $D\colon 2^X \to \mathbb{R}$ which measures the diversity of a given set of search points. Considering evolutionary diversity optimization, the goal is to find a set $P = \{I_1, \dots, I_\mu\}$ of μ solutions maximizing D among all sets of μ solutions under the condition that $q(I) \geq \alpha$ holds for all $I \in P$, where α is a given quality threshold. Here μ is the size of the set that we are aiming for, which determines the parent population size in our evolutionary diversity optimization approach.

We propose to use a very well-established concept, i.e., the use of indicators from multi-objective optimization. In multi-objective optimization, a function $g\colon X \to \mathbb{R}^d$ containing d objectives is given and all objectives should be optimized at the same time. As the given objectives are usually conflicting, one is interested in the trade-offs with respect to the given objective functions. Indicators in the area of multi-objective optimization have been used for many years to compare sets of solutions in the objective space, either for the purposes of comparing algorithm performance, or for use within an algorithm to drive a diversified search. Similarly to the diversity measure D in evolutionary diversity optimization, an indicator $I\colon 2^X \to \mathbb{R}$ measures the quality of a set of solutions according to some indicator function I. The immediate problem with applying

multi-objective optimization indicators is that diversity lacks the notion of dominance. In the context of multi-objective optimization, the optimal solutions are also referred to as non-dominated solutions. A solution x is called non-dominated (or Pareto optimal) if there is no other solution that is at least as good as x with respect to every objective and better in at least one objective. As multi-objective approaches aim to compute a set of non-dominated solutions, they reject dominated solutions over time. In evolutionary diversity optimization, every solution meeting the quality criteria is eligible and only the diversity among such solutions matters. Hence, we have to adapt the multi-objective indicators in a way that makes all solutions meeting the quality criterion non-dominated. We do this by ensuring that all solutions are incomparable when applying these indicators. For a more comprehensive introduction to dominance we refer the interested reader to [5], which is present in a large number of multi-objective optimization indicators.

In the following sections, we will first present existing multi-objective optimization indicators and our transformations to deal with the dominance issue. Then, we introduce the generic $(\mu + \lambda)$-EA_D and the concrete variants that will form the basis for our subsequent experimental studies on diversity optimization.

2.1 Multi-objective optimization indicators for diversity optimization

In this article, we use three quality indicators evaluating the quality of a given set of objective vectors S. For a given set of search points P (called the population) and a function $g\colon X \to \mathbb{R}^d$, we define $S = \{g(x) \mid x \in P\}$ as the set of objective vectors of P.

- **Hypervolume (HYP):** HYP is the volume covered by the set of objective vectors S with respect to a given reference point r. The hypervolume indicator measures the volume of the dominated space of all solutions contained in a set $S \subseteq \mathbb{R}^d$. This space is measured with respect to a given reference point $r = (r_1, r_2, \dots, r_d)$. The hypervolume $HYP(S, r)$ of a given set of objective vectors S with respect to r is then defined as $HYP(S, r) = VOL\left(\cup_{(s_1, \dots, s_d) \in S} [r_1, s_1] \times \cdots [r_d, s_d]\right)$ with $VOL(\cdot)$ being the Lebesgue measure.

- **Inverted generational distance (IGD):** IGD measures S with respect to a given reference set R. It calculates the average distance of objective vectors in R to their closest points in S. We have $IGD(R, S) = \frac{1}{|R|} \sum_{r \in R} \min_{s \in S} d(r, s)$, where $d(r, s)$ is the Euclidean distance between r and s in the objective space.

- **Additive epsilon approximation (EPS):** EPS measures the approximation quality of the worst approximated point in R that S achieves. For finite sets $S, R \subset \mathbb{R}^d$, the additive approximation of S with respect to R (assuming all objectives are to be minimized) is defined as $\alpha(R, S) := \max_{r \in R} \min_{s \in S} \max_{1 \leq i \leq d} (s_i - r_i)$. To get a sensitive indicator that can be used to guide the search, we consider instead the set $\{\alpha(\{r\}, S) \mid r \in R\}$ of all approximations of the points in R. We sort this set decreasingly and call the resulting sequence $S_\alpha(R, S) := (\alpha_1, \dots, \alpha_{|R|})$ (see [4, 32]).

While other indicators could also be used for driving diversity optimization, we do not intend to highlight differences of the indicators as this has been subject to many papers. Instead, we will focus on demonstrating that they can in-fact be used as a tool out-of-the-box to explore the space of combinations of instance features.
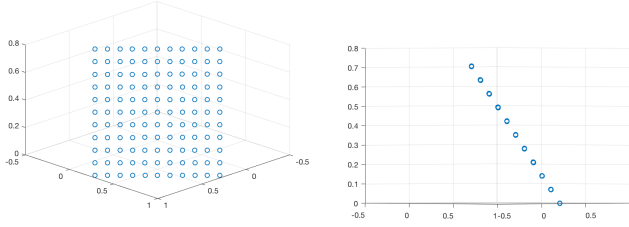
Figure 1: Reference set in 3D using 11^2 objective vectors. The normal vector that goes through the centre of the square goes through the origin. We use 101^2 feature vectors in our experiments. Note that all solution vectors are by definition on the Pareto front, i.e., in the unit square.

These three indicators cannot be applied immediately, as there is no reference set (which some indicators require) and one has to deal with the issue of dominance as there is no preference of one feature value over the other. For example, let us consider two scaled features and visualize the combinations as points in a two-dimensional unit square. In this case, we would like to cover the entire square evenly, without preferring one region over the other, and in particular we cannot say that one area is preferred over another – a naive multi-objective optimization setup for this two-dimensional problem might focus, for example, only on the area near the origin.

We propose two approaches to deal with this challenge: (1) transformation of the two-dimensional problem into a three-dimensional problem, (2) doubling the number of dimensions.

2.1.1 Problem Transformation. When we are interested in covering a two-dimensional feature space, we can mitigate the problem of EPS-/HYP-preferred regions by transforming the two-dimensional problem into a three-dimensional one. We do so as follows:

(1) We place the unit square with its original x/y-coordinates in the three-dimensional space using $z = 0$.
(2) We rotate it around the x and y axis by 45 degrees each time.
(3) We translate it such that the center point of the transformed unit square is at $(sqrt(2)/4)^3$ (see Figure 1).

After these steps, the normal vector that goes through the center of the unit square also goes through the origin. This orientation allows us to use the wide spectrum of well-established quality indicators from the field of multi-objective optimization, designed for assessing various aspects of solutions sets, such as convergence and distribution – and no modifications are needed at all. Especially for the volume- and dominance-based indicators our transformation has the important benefit that all features are of equal importance.

As we perform the same transformation with the instance set (i.e., our population) as well as the reference set (after rescaling it into the unit square based on known lower and upper values for the features), this means that the population is always on the Pareto front; this is a situation that is not that common in multi-objective optimization. Our goal is now to cover the reference set "evenly", as defined by the respective indicators.

2.1.2 Dimension doubling. To avoid the dominance issue, we propose the following transformation. Given a feature vector $p = (p_1, p_2, \ldots, p_d)$ in the d-dimensional space, we project it into the $2d$-dimensional space by copying the original feature values and negating their copy, resulting in $p' = (p_1, p_2, \ldots, p_d, -p_1, -p_2, \ldots, -p_d)$. With this, dominance between solution vectors vanishes, and we

---

Algorithm 1: $(\mu + \lambda)$-EA_D

---

1 Initialize the population P with μ instances of quality at least α.
2 Let $C \subseteq P$ where $|C| = \lambda$.
3 For each $I \in C$, produce an offspring I' of I by mutation. If $q(I') \geqslant \alpha$, add I' to P.
4 While $|P| > \mu$, remove an individual with the smallest loss to the diversity indicator D.
5 Repeat step 2 to 4 until termination criterion is reached.

---

can employ the hypervolume indicator without the need for any modifications.

Because we work with rescaled value ranges in $[0, 1]^d$, the necessary hypervolume reference point r has to be adequately chosen in the 2d-dimensional space. For example. $(1^d, 0^d)$ would be based on the ranges' extreme values, and $(2^d, 1^d)$ would put an increased focus on maintaining extreme points in the population.

While this transformation mitigates the dominance issue, it remains an open problem how this can be made to work with the epsilon indicator as well. The challenge here is to define an evenly spread out reference set in the 2d-dimensional space given our dimension doubling.

2.2 Evolutionary algorithm for optimizing diversity

The algorithm used to optimize the feature-based population diversity follows the setting in [13] with modifications. Algorithm 1 shows the evolutionary algorithm used for optimizing diversity. Let $I \in P$ be an individual in a population P. A problem specific feature vector $f(I) = (f_1(I), \ldots, f_d(I))$ is used to describe a potential solution. The indicators are calculated based on the feature vector.

Since the indicators introduced are defined in the space of $[0, 1]^d$, the feature values are scaled before the calculation of indicators. Let $f_i^{\max}$ and $f_i^{\min}$ be the maximum and minimum value of a certain feature f_i obtained from some initial experiments. The feature values are normalized based on the formula $f_i'(I) = (f_i(I) - f_i^{\min})/(f_i^{\max} - f_i^{\min})$. Feature values outside the range $[f_i^{\min}, f_i^{\max}]$ are set to 0 or 1, to allow the algorithm to work with non-anticipated features values.

Based on this, we investigate the following diversity-optimizing algorithms in this study:

- $EA_{\text{HYP-2D}}$ and EA_{EPS} use the idea of transforming the two-dimensional problem into a three-dimensional one.
- EA_{HYP} uses the idea of doubling the dimensions.
- EA_{IGD} uses IGD, which can be used without the need to transform the feature vectors, as it does not consider concepts like dominance or volume like HYP and EPS.

In addition, we use EA_{DIS} with discrepancy minimization, as used in [23]. As IGD and EPS require a reference set (e.g. solutions situated on the Pareto front), we use regular grids in the unit square and unit cube with a resolution of 101^2 solutions and 11^3 solutions. The necessary hypervolume reference point r for $EA_{\text{HYP-2D}}$ is set based on the extreme values of the reference set after the described rotations; for EA_{HYP} it is set to $(2^d, 1^d)$ to increase the focus on extreme points.

Note that these indicators are a major differentiator from the work in [18]. There, the approach was able of only evolving one instance at a time with the goal of reaching a particular target vector.

Figure 2: Image I^*.

	Notation	f^{min}	f^{max}	Description
f_1	SDHue	0.420	0.700	standard deviation hue
f_2	saturation	0.420	0.500	mean saturation
f_3	symmetry	0.715	0.740	reflectional symmetry
f_4	hue	0.250	0.400	color descriptor
f_5	GCF	0.024	0.027	Global Contrast Factor
f_6	smoothness	0.906	0.918	smoothness

Table 1: Description of features for images.

In our case, we evolve an entire population of diverse instances and do not require explicitly set targets.

3 IMAGES

In this section, we aim to evolve a diverse set of images as described in [1]. Given an image I^*, we want to compute a diverse set of images $P = \{I_1, \ldots, I_\mu\}$ that agree on a given quality criteria $q(I)$ for each $I \in P$. We will use the image I^* given in Figure 2 for our investigations. An image I fulfills the quality criteria $q(I)$ if the *mean-squared error* in terms of the RGB-values of I with respect to I^* is less than 500. Many different features have been widely applied to measurements of the properties of images. They often provide a good characterization of images. We select the set of features identified in [1]. We carry out the indicator-based evolutionary optimization approach with respect to different multi-objective indicators and different sets of features. Our evolutionary algorithm evolves diverse populations of images for each indicator and for each feature combination.

In our experiments we used the following features: standard-deviation-hue, mean-saturation, reflectional symmetry [11], hue [14], Global Contrast Factor [16], and smoothness [25]. Instead of applying the star discrepancy [29] to measure diversity, we use the multi-objective indicators as previously introduced. Otherwise, the configuration of Algorithm 1 is the same as in [23]. In order to produce a new solution the algorithm uses a self-adaptive offset random walk mutation introduced in [23]. Based on a random walk on the image this operator alters the RGB-values of the pixels visited in a slight way such that a new but similar image is obtained. Random walk lengths are increased in the case of a successful mutation and decreased in the case of unsuccessful ones. For details, we refer the reader to [23, 24].

3.1 Experimental settings

Now, we consider the indicator-based diversity optimization for combinations of two and three features. We select features in order to combine different aesthetic and general features based on our initial experimental investigations and previous investigations

in [22]. In this work we explore several features and features ranges described in Table 1. We use scaled feature values while we calculate the different indicators values of a given set of points. After having consider the combination of two features, we investigate sets of three features. Here, we select different features combining aesthetic and general features together used in the previous experiment.

In order to obtain a clear comparison between our present experiments and experiments based on the discrepancy-based evolutionary algorithm introduced in [23] we work with the same range of feature values.

We run each configuration for 2000 generations with a population size of $\mu = 20$ and $\lambda = 1$. To assess our results using statistical tests, we run each combination of feature-pair and indicator 30 times. All algorithms were implemented in *Matlab* ($R2017b$) and run on 48-core compute nodes with AMD 2.80 GHz CPUs and 128 GB of RAM.

3.2 Experimental results and analysis

We present a series of experiments for two- and three-feature combinations in order to evaluate our evolutionary diversity algorithms based on the use of indicators from multi-objective optimization described in Section 2.

3.2.1 Two-feature combinations. Our results are summarized in Table 2 and Table 3. The columns represent the algorithms with the corresponding mean value and standard deviation. The rows represent the indicators HYP-2D, HYP, IGD, EPS and discrepancy (DIS). For each indicator, we obtained results for all sets of features.

Additionally, we use the Kruskal-Wallis test for statistical validation with 95% confidence and subsequently apply the Bonferroni post-hoc statistical procedure. For a detailed description of the statistical tests we refer the reader to [7]. Our experimental analysis characterizes the behavior of the four examined indicator-based evolutionary algorithms and discrepancy-based evolutionary algorithm. In the statistical tests shown in Table 2 and Table 3, $A^{(+)}$ is equivalent to the statement that the algorithm in this column outperformed algorithm A, and $A^{(-)}$ is equivalent to the statement that A outperformed the algorithm given in the column. If the algorithm A does not appear, this means that no significant difference was determined.

Figure 3 illustrates feature plots of (randomly selected) final populations of EA_{HYP-2D} (top), EA_{HYP}, EA_{IGD} and EA_{EPS} (bottom) for three pairs of feature combinations. In the first column, we see the feature vectors for the final population of the four algorithms for image based on pairs of features (f_1, f_2). It can be observed that the discrepancy value for EA_{HYP-2D} is 0.1767. This is significantly smaller than the one for EA_{EPS} at 0.6802. Note that smaller discrepancy values are considered to be better. The middle column shows the combination of the feature pair (f_3, f_4). The discrepancy value for feature pair (f_3, f_4) for EA_{IGD} is 0.2286 whereas it is 0.6015 for EA_{EPS}. The last column shows the final populations of the diversity optimization when considering feature pair (f_5, f_6). The discrepancy value for feature pair (f_5, f_6) is the smallest among all algorithms for EA_{HYP-2D} at 0.2182 and the highest for EA_{EPS} at 0.6318.

In summary, we observe that EA_{HYP-2D}, EA_{HYP} and EA_{IGD} achieve a good and even coverage of the feature space, especially in comparison with the discrepancy-based diversification (see Figure 4 for an example from [23]). Interestingly, EA_{EPS} appears to experience difficulties, and it achieves the worst coverage in the search space in all scenarios.
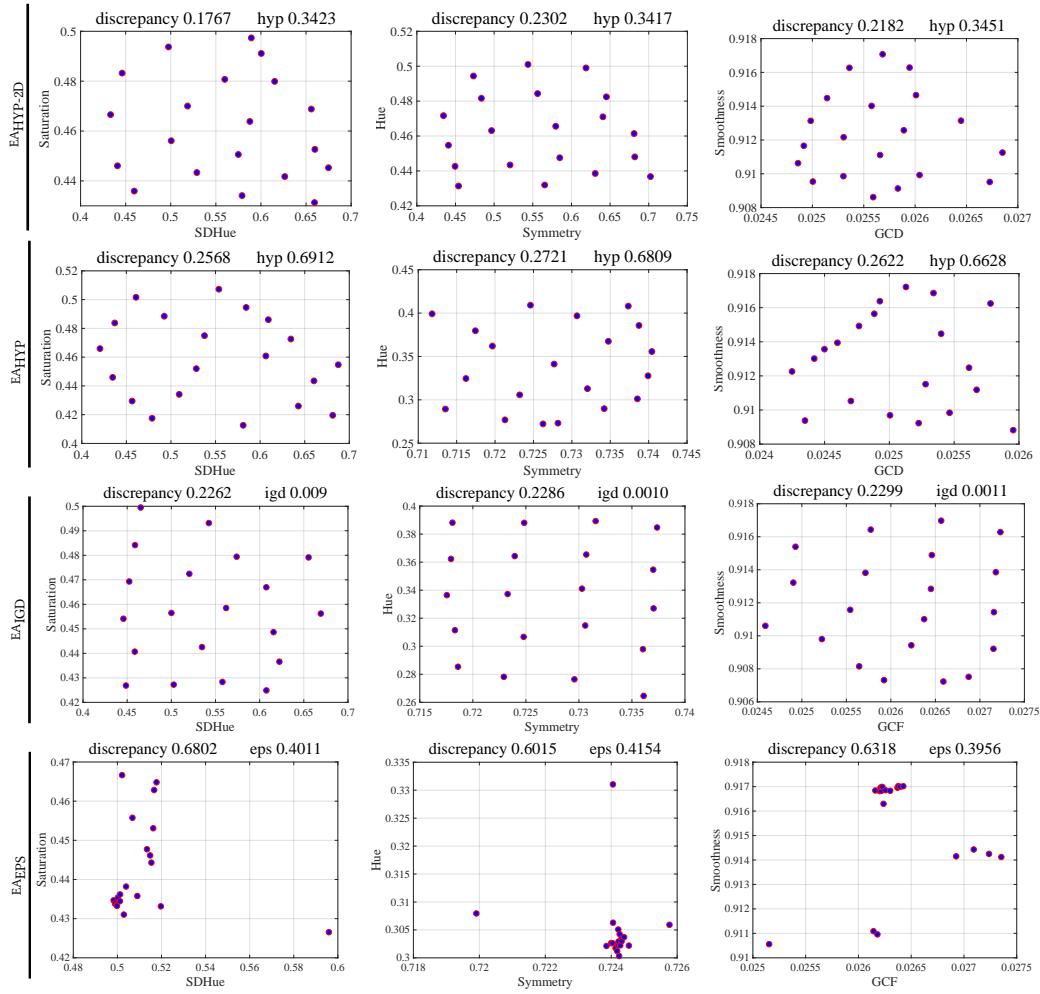
Figure 3: Feature vectors for final population of EA_HYP-2D (top), EA_HYP, EA_IGD and EA_EPS (bottom) for images based on pair of features from left to right: (f_1, f_2), (f_3, f_4), (f_5, f_6).

Table 2: Investigations for images with 2 features. Comparison in terms of mean, standard deviation and statistical test for considered indicators.

|		EA_HYP-2D (1)			EA_HYP (2)			EA_IGD (3)			EA_EPS (4)			EA_DIS (5)		
		mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat
HYP-2D	$f_1.f_2$	0.347	0.004	$4^{(+)},5^{(+)}$	0.382	0.007	$3^{(+)},4^{(+)},5^{(+)}$	0.335	0.003	$2^{(-)},5^{(+)}$	0.198	0.019	$1^{(-)},2^{(-)}$	0.112	0.030	$1^{(-)},2^{(-)},3^{(-)}$
	$f_3.f_4$	0.344	0.004	$2^{(+)},4^{(+)},5^{(+)}$	0.268	0.014	$1^{(-)},3^{(-)},4^{(-)},5^{(+)}$	0.339	0.004	$2^{(+)},4^{(+)},5^{(+)}$	0.221	0.015	$1^{(-)},2^{(-)},3^{(-)}$	0.105	0.025	$1^{(-)},2^{(-)},3^{(-)}$
	$f_5.f_6$	0.350	0.007	$2^{(+)},3^{(+)},4^{(+)},5^{(+)}$	0.342	0.004	$1^{(-)},4^{(+)},5^{(+)}$	0.332	0.004	$1^{(-)},4^{(+)},5^{(+)}$	0.220	0.045	$1^{(-)},2^{(-)},3^{(-)}$	0.134	0.016	$1^{(-)},2^{(-)},3^{(-)}$
HYP	$f_1.f_2$	0.525	0.012	$3^{(+)},4^{(+)},5^{(+)}$	0.693	0.013	$3^{(+)},4^{(+)},5^{(+)}$	0.374	0.006	$1^{(-)},2^{(-)},4^{(+)}$	0.344	0.003	$1^{(-)},2^{(-)},3^{(-)}$	0.363	0.014	$1^{(-)},2^{(-)}$
	$f_3.f_4$	0.500	0.007	$3^{(+)},4^{(+)},5^{(+)}$	0.681	0.010	$3^{(+)},4^{(+)},5^{(+)}$	0.268	0.072	$1^{(-)},2^{(-)},4^{(+)},5^{(+)}$	0.280	0.010	$1^{(-)},2^{(-)},3^{(-)}$	0.267	0.014	$1^{(-)},2^{(-)},3^{(-)}$
	$f_5.f_6$	0.518	0.012	$2^{(-)},4^{(+)},5^{(+)}$	0.663	0.010	$1^{(+)},3^{(+)},4^{(+)},5^{(+)}$	0.335	0.004	$2^{(-)},4^{(+)}$	0.317	0.006	$1^{(-)},2^{(-)},3^{(-)}$	0.327	0.008	$1^{(-)},2^{(-)}$
IGD	$f_1.f_2$	0.001	0.335	$2^{(+)},4^{(+)},5^{(+)}$	0.003	0.000	$1^{(-)},3^{(-)}$	0.001	0.000	$2^{(+)},4^{(+)},5^{(+)}$	0.003	0.000	$1^{(-)},3^{(-)},5^{(+)}$	0.005	0.001	$1^{(-)},3^{(-)},4^{(-)}$
	$f_3.f_4$	0.001	0.339	$2^{(+)},4^{(+)},5^{(+)}$	0.004	0.000	$1^{(-)},3^{(-)},5^{(+)}$	0.001	0.000	$2^{(+)},4^{(+)},5^{(+)}$	0.003	0.000	$1^{(-)},3^{(-)},5^{(+)}$	0.005	0.001	$1^{(-)},2^{(-)},3^{(-)},4^{(-)}$
	$f_5.f_6$	0.002	0.332	$2^{(+)},5^{(+)}$	0.007	0.000	$1^{(-)},3^{(-)},4^{(-)},5^{(-)}$	0.001	0.000	$2^{(+)},4^{(+)},5^{(+)}$	0.003	0.001	$2^{(+)},3^{(-)}$	0.004	0.001	$1^{(-)},2^{(+)},3^{(-)}$
EPS	$f_1.f_2$	0.190	0.198	$2^{(+)},4^{(+)},5^{(+)}$	0.498	0.011	$1^{(-)},3^{(-)}$	0.194	0.032	$2^{(+)},4^{(+)},5^{(+)}$	0.402	0.039	$1^{(-)},3^{(-)},5^{(+)}$	0.600	0.106	$1^{(-)},3^{(-)},4^{(-)}$
	$f_3.f_4$	0.198	0.221	$2^{(+)},4^{(+)},5^{(+)}$	0.569	0.016	$1^{(-)},3^{(-)}$	0.208	0.035	$2^{(+)},4^{(+)},5^{(+)}$	0.418	0.036	$1^{(-)},3^{(-)},5^{(+)}$	0.615	0.069	$1^{(-)},3^{(-)},4^{(-)}$
	$f_5.f_6$	0.125	0.220	$2^{(+)},4^{(+)},5^{(+)}$	0.946	0.001	$1^{(-)},3^{(-)},4^{(-)}$	0.225	0.064	$2^{(+)},4^{(+)},5^{(+)}$	0.397	0.110	$1^{(-)},2^{(-)},3^{(-)}$	0.587	0.063	$1^{(-)},3^{(-)}$
DIS	$f_1.f_2$	0.171	0.018	$2^{(+)},4^{(+)},5^{(+)}$	0.257	0.010	$1^{(-)},4^{(+)}$	0.201	0.031	$4^{(+)},5^{(+)}$	0.686	0.064	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.204	0.116	$1^{(-)},3^{(-)},4^{(+)}$
	$f_3.f_4$	0.234	0.031	$4^{(+)}$	0.273	0.041	$3^{(-)},4^{(+)},5^{(-)}$	0.198	0.017	$2^{(+)},4^{(+)}$	0.606	0.054	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.228	0.059	$2^{(+)},4^{(+)}$
	$f_5.f_6$	0.221	0.026	$4^{(+)}$	0.263	0.070	$3^{(-)},4^{(+)},5^{(-)}$	0.205	0.055	$2^{(+)},4^{(+)}$	0.633	0.158	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.203	0.054	$2^{(+)},4^{(+)}$
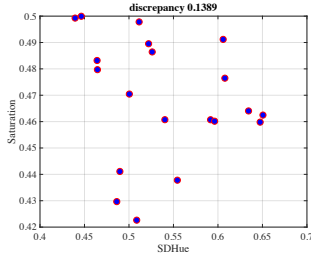
Figure 4: Feature vectors for final population of EA$_{DIS}$ [23] for images based on (f_1, f_2).

Table 3: Investigations for images with 3 features. Comparison in terms of mean, standard deviation and statistical test for considered indicators.

| | | EA$_{HYP}$ (1) | | | EA$_{IGD}$ (2) | | | EA$_{DIS}$ (3) | |
		mean	st	stat	mean	st	stat	mean	st	stat
HYP	f_1,f_2,f_3	0.5251	0.0122	$2^{(+)},3^{(+)}$	0.2096	0.0018	$1^{(-)},3^{(-)}$	0.2196	0.0110	$1^{(-)},2^{(+)}$
	f_1,f_4,f_3	0.4998	0.0071	$2^{(+)},3^{(+)}$	0.2142	0.0036	$1^{(-)},3^{(-)}$	0.2286	0.0034	$1^{(-)},2^{(+)}$
	f_5,f_4,f_2	0.5181	0.0122	$2^{(+)},3^{(+)}$	0.1785	0.0017	$1^{(-)},3^{(-)}$	0.1961	0.0023	$1^{(-)},2^{(+)}$
IGD	f_1,f_2,f_3	0.0146	0.0001	$2^{(-)},3^{(+)}$	0.0067	0.0003	$1^{(+)},3^{(+)}$	0.0148	0.0003	$1^{(-)},2^{(-)}$
	f_1,f_4,f_3	0.0150	0.0001	$2^{(-)}$	0.0074	0.0002	$1^{(+)},3^{(+)}$	0.0151	0.0001	$2^{(-)}$
	f_5,f_4,f_2	0.0193	0.0001	$2^{(-)},3^{(+)}$	0.0062	0.0002	$1^{(+)},3^{(+)}$	0.0199	0.0007	$1^{(-)},2^{(-)}$
DIS	f_1,f_2,f_3	0.3554	0.0458	$2^{(+)},3^{(-)}$	0.3809	0.0522	$1^{(-)},3^{(-)}$	0.3350	0.1002	$1^{(+)},2^{(+)}$
	f_1,f_4,f_3	0.3493	0.0532	$2^{(-)}$	0.2860	0.0342	$1^{(+)},3^{(+)}$	0.3118	0.1309	$2^{(-)}$
	f_5,f_4,f_2	0.4237	0.0643	$2^{(-)},3^{(-)}$	0.3227	0.0557	$1^{(+)},3^{(-)}$	0.3007	0.1467	$1^{(+)},2^{(-)}$

Moreover, in Table 2, we observe that the EA$_{HYP}$ algorithm has the best performance among all algorithms. It has the highest hypervolume values for all features combinations, and this is also statistically significant. Also, due to the statistical tests we can say that EA$_{HYP-2D}$ outperforms EA$_{EPS}$ and EA$_{DIS}$ with respect to the inverted generational distance and additive epsilon approximation indicator measurements values for all sets of features. We observe that EA$_{HYP-2D}$ considering IGD and EPS values has no significant differences to EA$_{IGD}$. In terms of discrepancy, the EA$_{HYP-2D}$ has a following characteristic: for set of features (f_1,f_2) the EA$_{HYP-2D}$ outperforms EA$_{HYP}$, EA$_{EPS}$ and EA$_{DIS}$, however, it only outperforms the EA$_{EPS}$ for the set of features (f_3,f_4) and (f_5,f_6).

Furthermore, EA$_{IGD}$ outperforms the EA$_{HYP}$, EA$_{EPS}$ and the EA$_{DIS}$ with respect to IGD, EPS and DIS indicators measurements in most of the cases and achieves the lowest values for IGD measurements among all others algorithms for all sets of features. The best performance achieves EA$_{IGD}$ for discrepancy measurements for the combinations of features (f_3,f_4) and (f_5,f_6) with values 0.198 and 0.205. The hypervolume-based approaches EA$_{HYP-2D}$ and EA$_{HYP}$ outperform EA$_{IGD}$ for all sets of features. Among all other algorithms EA$_{EPS}$ shows the worst performance. Especially, according to all indicators measurements and all sets of features, the EA$_{EPS}$ is dominated by EA$_{HYP}$ and EA$_{IGD}$, and this difference is statistically significant. Finally, the EA$_{DIS}$ is dominated by EA$_{HYP-2D}$ and EA$_{HYP}$, EA$_{IGD}$ and EA$_{EPS}$ with respect to the HYP-2D, HYP, IGD and EPS indicator values. Also, most results are significantly different with respect to the HYP, IGD, EPS indicators. EA$_{DIS}$ achieves the best performance with respect to the DIS indicator for the combinations of features (f_3,f_4) and (f_5,f_6). The EA$_{DIS}$ outperforms the EA$_{HYP}$ and EA$_{EPS}$ in this case. For the combinations (f_1,f_2) with respect to the DIS indicator, the EA$_{DIS}$ is dominated by EA$_{HYP-2D}$ and EA$_{IGD}$.

3.2.2 Three-feature combinations. The triplets of features are described in Table 1 and the results are summarized in Table 3. As

Table 4: Description of features for TSP instances.

	Notation	f^{min}	f^{max}
f_1	angle_mean	0.70	2.90
f_2	centroid_mean_distance_to_centroid	0.24	0.70
f_3	nnds_mean	0.10	0.70
f_4	mst_dists_mean	0.06	0.15

before, the columns represent the algorithms with the corresponding mean value and standard deviation, and the rows represent the indicators. In Table 3, we compare EA$_{HYP}$ and EA$_{IGD}$ with EA$_{DIS}$ algorithm with respect to two multi-objective indicators and the discrepancy measurement. Table 3 shows that EA$_{HYP}$ outperforms EA$_{IGD}$ and EA$_{DIS}$ for all three sets of features with respect to the HYP indicator. In particular, for the first set of features (f_1,f_2,f_3) the EA$_{HYP}$ algorithm obtains the value 0.5251, and only 0.2096 for IGD, and 0.2196 for discrepancy.

Comparing EA$_{IGD}$ to EA$_{HYP}$ and EA$_{DIS}$ with respect to the IGD indicator, we find a similar picture as for the EA$_{HYP}$ algorithm. EA$_{IGD}$ clearly outperforms the EA$_{HYP}$ and EA$_{DIS}$ for all three sets of features. The EA$_{DIS}$ algorithm also clearly outperforms EA$_{HYP}$ and EA$_{IGD}$ with respect to discrepancy. Overall, EA$_{DIS}$ achieves improvements in terms of discrepancy value among another two algorithms for all sets of features apart from one exception. It can be observed that for the set of features (f_1,f_4,f_3) EA$_{DIS}$ does not have a major advantage over the EA$_{IGD}$.

In a nutshell, according to our statistical tests the EA$_{HYP}$ outperforms all examined algorithms with respect to the HYP indicator values for all sets of features in case of two-feature combination. Moreover, EA$_{IGD}$ outperforms EA$_{HYP}$, EA$_{EPS}$ and EA$_{DIS}$ with respect to the IGD indicator, which was expected, but it shows no significant difference to EA$_{HYP-2D}$ for the first two sets of features. The EA$_{EPS}$ algorithm has the worst performance, no matter the indicator considered. Similarly, considering our experiments for three-feature combinations, EA$_{HYP}$ and EA$_{IGD}$ achieve the best results, which are also statistically significant.

4 TRAVELING SALESPERSON PROBLEM

We now investigate our approaches for evolving diverse sets of instances for the Euclidean Traveling Salesperson problem (TSP) building on the work carried out in [13, 17, 19–21].

Our goal is to generate diverse sets of TSP instances with 50 cities in the space of $[0, 1]^2$, which is a reasonable size of problem for feature analysis of TSP. The instance quality is evaluated by the approximation ratio, which is given by $\alpha_A(I) = A(I)/OPT(I)$, where $A(I)$ is the fitness value of the solution found by algorithm A for the given instance I, and $OPT(I)$ is the size of an optimal solution for instance I which in our case is calculated using the exact TSP solver Concorde [2]. Within this study, $A(I)$ is the minimum tour length obtained by three independent repeated runs of the 2-OPT algorithm for a given TSP instance I. As the number of cities in an instance is 50, our algorithm chooses 1.18 as threshold for approximation ratio, which means only TSP instances with approximation ratios equal to or greater than 1.18 are accepted; this follows the setting in [13].

4.1 Experimental settings

The algorithm is implemented in R [26], and the feature vectors are calculated using the tspmeta package [17]. The hardware is identical to that used in the image-related experiments. The features
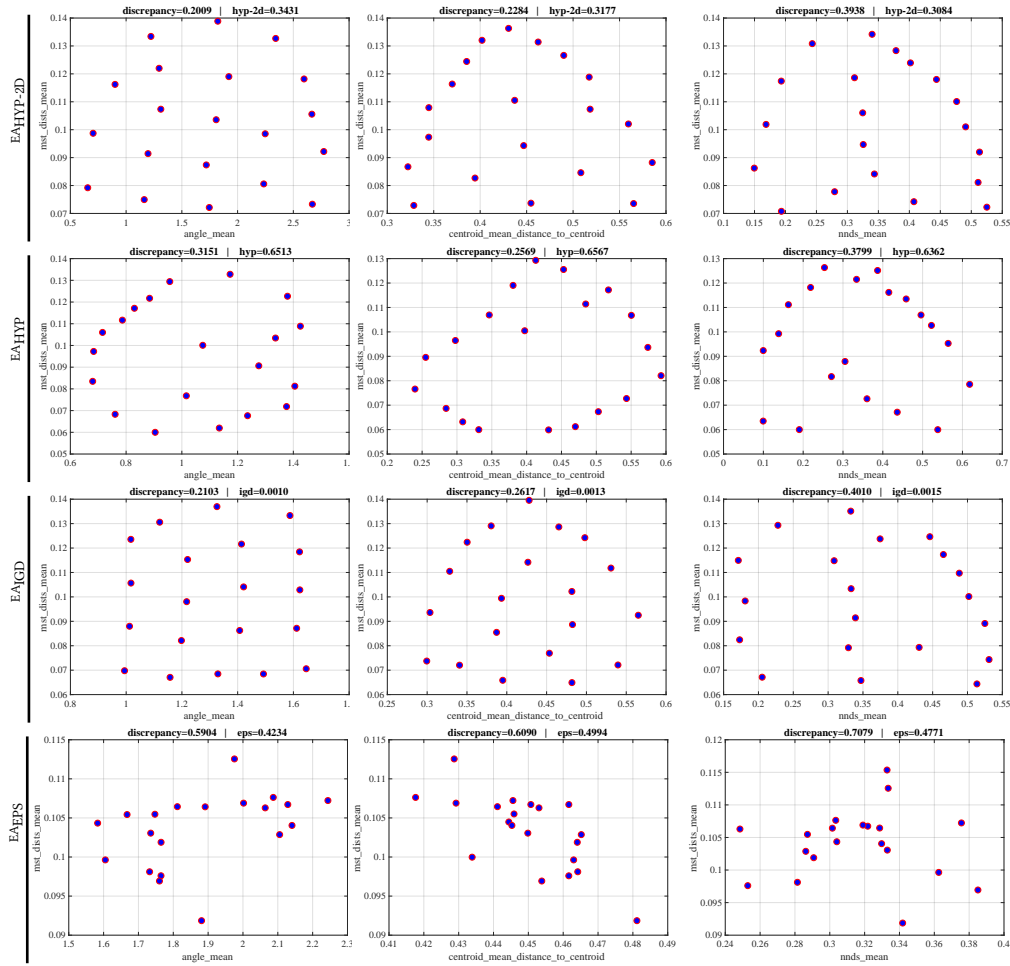
Figure 5: Feature vectors for final population of EA$_{HYP-2D}$ (top), EA$_{HYP}$ (2nd), EA$_{IGD}$ (3rd) and EA$_{EPS}$ (bottom) for TSP instances based on two features from left to right: (f_1, f_4), (f_2, f_4), (f_3, f_4).

we use to characterize TSP instances are given in Table 4. A detailed explanation of the features can be found in [17]. The parameter setting follows the same setting as in [23]. The population size μ and number of offspring generated λ of EA is set to 20 and 1 respectively. As mentioned before in Section 2, we normalize feature values before indicator calculations. Based on the results gathered from some initial runs of feature-based diversity maximization algorithm, the maximum and minimum values f^{max} and f^{min} for each feature are determined (see Table 4). Each algorithm setting is repeated independently for 30 times. Each experiment is run for 20000 generations and the values of all proposed indicators and discrepancy values are reported in the following section.

4.2 Experimental results and analysis

As before, three pairs of features and three triplets of features are examined. The results are compared with those from the discrepancy minimization algorithm.

4.2.1 Two-feature combinations. Figure 5 shows some (randomly drawn) populations in the feature space after running the

corresponding algorithms with consideration of certain two-feature combinations. In these figures, the populations after optimizing the hypervolume and inverted generational distance show good coverage and distribution over the whole space. Compared to the 2D plots from previous research [12, 23], the EA maximizing HYP-2D is able to generate individuals with feature vectors that are not found in previous research. The feature vectors obtained from EA$_{HYP-2D}$, EA$_{HYP}$ and EA$_{IGD}$ are – in our opinion – nicely distributed in the space. In respect of indicator values, the population discrepancies of the sample populations from EA$_{HYP-2D}$, EA$_{HYP}$ and EA$_{IGD}$ are comparable to those from the algorithm minimizing discrepancy value. Although the discrepancy values are similar, the individuals from these three algorithms are better distributed than the previous results [23].

Table 5 lists the results of 30 independent runs, following the same layout as Table 2. The statistics are gathered from the final populations after running each algorithm on the three different two-feature combinations. The statistical values in the first three large columns are from the EA maximizing HYP-2D, HYP and minimizing IGD respectively. The results show that they outperform

Table 5: Investigations for TSP instances with 2 features. Comparison in terms of mean, standard deviation and statistical test for considered indicators.

		EA_{HYP-2D} (1)			EA_{HYP} (2)			EA_{IGD} (3)			EA_{EPS} (4)			EA_{DIS} (5)		
		mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat
HYP-2D	f_1,f_4	0.338	2E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.309	4E-3	$1^{(-)},4^{(+)}$	0.331	3E-3	$4^{(+)},5^{(+)}$	0.190	1E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.256	1E-2	$1^{(-)},3^{(-)}$
	f_2,f_4	0.317	3E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.303	5E-3	$1^{(-)},3^{(-)},4^{(+)}$	0.316	3E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.178	1E-7	$1^{(-)},2^{(-)},3^{(-)}$	0.252	1E-2	$1^{(-)},3^{(-)}$
	f_3,f_4	0.303	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.296	5E-3	$1^{(-)},3^{(-)},4^{(+)},5^{(+)}$	0.304	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.190	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.238	2E-2	$1^{(-)},2^{(-)},3^{(-)}$
HYP	f_1,f_4	0.645	5E-3	$4^{(+)},5^{(+)}$	0.638	7E-3	$4^{(+)},5^{(+)}$	0.639	6E-3	$4^{(+)},5^{(+)}$	0.424	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.529	3E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_2,f_4	0.609	7E-3	$2^{(-)},4^{(+)},5^{(+)}$	0.632	1E-2	$1^{(-)},4^{(+)},5^{(+)}$	0.621	6E-3	$4^{(+)},5^{(+)}$	0.398	1E-6	$1^{(-)},2^{(-)},3^{(-)}$	0.505	2E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_3,f_4	0.584	3E-2	$2^{(-)},4^{(+)}$	0.621	9E-3	$1^{(-)},3^{(+)},4^{(+)},5^{(+)}$	0.595	4E-2	$2^{(-)},4^{(+)},5^{(+)}$	0.410	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.485	3E-2	$2^{(-)},3^{(-)}$
IGD	f_1,f_4	0.001	2E-5	$4^{(+)},5^{(+)}$	0.001	6E-5	$3^{(-)},4^{(+)}$	0.001	4E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.003	2E-5	$1^{(-)},2^{(-)},3^{(-)}$	0.002	2E-4	$1^{(-)},3^{(-)}$
	f_2,f_4	0.001	3E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.002	6E-5	$1^{(-)},3^{(-)},4^{(+)}$	0.001	3E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.003	2E-10	$1^{(-)},2^{(-)},3^{(-)}$	0.002	2E-4	$1^{(-)},3^{(-)}$
	f_3,f_4	0.002	3E-4	$4^{(+)},5^{(+)}$	0.002	6E-5	$3^{(-)},4^{(+)},5^{(+)}$	0.002	3E-4	$2^{(+)},4^{(+)},5^{(+)}$	0.003	3E-5	$1^{(-)},2^{(-)},3^{(-)}$	0.003	3E-4	$1^{(-)},2^{(-)},3^{(-)}$
EPS	f_1,f_4	0.196	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.249	2E-2	$1^{(-)},3^{(-)},4^{(+)}$	0.189	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.423	1E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.345	4E-2	$1^{(-)},3^{(-)}$
	f_2,f_4	0.226	8E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.256	2E-2	$1^{(-)},3^{(-)},4^{(+)},5^{(+)}$	0.228	1E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.499	2E-16	$1^{(-)},2^{(-)},3^{(-)}$	0.360	5E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_3,f_4	0.260	4E-2	$4^{(+)},5^{(+)}$	0.278	2E-2	$4^{(+)},5^{(+)}$	0.265	4E-2	$4^{(+)},5^{(+)}$	0.477	3E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.368	5E-2	$1^{(-)},2^{(-)},3^{(-)}$
DIS	f_1,f_4	0.222	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.353	2E-2	$1^{(-)},3^{(-)},4^{(+)}$	0.249	2E-2	$2^{(+)},4^{(+)}$	0.589	4E-3	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.292	5E-2	$1^{(-)},4^{(+)}$
	f_2,f_4	0.230	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.274	2E-2	$1^{(-)},4^{(+)},5^{(+)}$	0.252	1E-3	$4^{(+)},5^{(+)}$	0.609	1E-16	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.336	4E-2	$1^{(-)},2^{(-)},3^{(-)},4^{(+)}$
	f_3,f_4	0.418	6E-2	$4^{(+)}$	0.416	3E-2	$4^{(+)}$	0.401	7E-2	$4^{(+)},5^{(+)}$	0.719	6E-3	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.448	9E-2	$3^{(-)},4^{(+)}$

Table 6: Investigations for TSP instances with 3 features. Comparison in terms of mean, standard deviation and statistical test for considered indicators.

		EA_{HYP} (1)			EA_{IGD} (2)			EA_{DIS} (3)		
		mean	st	stat	mean	st	stat	mean	st	stat
HYP	f_1,f_2,f_3	0.4511	1E-2	$2^{(+)},3^{(+)}$	0.4261	7E-3	$1^{(-)},3^{(+)}$	0.3385	6E-3	$1^{(-)},2^{(-)}$
	f_1,f_3,f_4	0.4579	8E-3	$2^{(+)},3^{(+)}$	0.4260	6E-3	$1^{(-)},3^{(+)}$	0.3430	6E-3	$1^{(-)},2^{(-)}$
	f_2,f_3,f_4	0.4478	8E-3	$2^{(+)},3^{(+)}$	0.4262	6E-3	$1^{(-)},3^{(+)}$	0.3430	6E-3	$1^{(-)},2^{(-)}$
IGD	f_1,f_2,f_3	0.0083	3E-4	$2^{(-)},3^{(+)}$	0.0075	2E-4	$1^{(+)},3^{(+)}$	0.0110	1E-4	$1^{(-)},2^{(-)}$
	f_1,f_3,f_4	0.0082	2E-4	$2^{(-)},3^{(+)}$	0.0077	1E-4	$2^{(+)},3^{(+)}$	0.0107	1E-4	$1^{(-)},2^{(-)}$
	f_2,f_3,f_4	0.0086	2E-4	$2^{(-)},3^{(+)}$	0.0080	2E-2	$2^{(+)},3^{(+)}$	0.0112	8E-5	$1^{(-)},2^{(-)}$
DIS	f_1,f_2,f_3	0.4115	3E-2	$2^{(+)},3^{(+)}$	0.4839	3E-2	$1^{(-)},3^{(-)}$	0.4399	2E-2	$1^{(-)},2^{(+)}$
	f_1,f_3,f_4	0.5220	4E-2	$3^{(-)}$	0.5474	3E-2	$3^{(-)}$	0.4757	2E-2	$1^{(+)},2^{(+)}$
	f_2,f_3,f_4	0.4669	3E-2	$2^{(+)}$	0.5111	3E-2	$1^{(-)},3^{(-)}$	0.4667	2E-2	$2^{(+)}$

the evolutionary algorithms minimizing EPS and discrepancy in all four indicators. Both EA_{HYP-2D} and EA_{IGD} achieve significant improvements in all four indicators after running for 20000 generations. It is not a surprise that EA_{HYP-2D} outperforms the other three algorithms in terms of hypervolume covered. It also shows comparable performance in optimizing IGD and other indicators. The same behavior is observed for EA_{IGD}, which outperforms EA_{EPS} and EA_{DIS} and maximizes HYP relatively well. EA_{DIS} is designed for the purpose of minimizing the population discrepancy value. However, based on the statistical analysis, it does not obtain better population discrepancy than EA_{HYP-2D} or EA_{IGD} after 20000 generations. Similar to what we have observed in the image-based study in Section 3, the results of EA_{EPS} are not as good as those from the algorithms optimizing HYP-2D, HYP and IGD. No significant improvement in population diversity is achieved using this algorithm. We have experimented with target grids of higher resolution to mitigate local-sensitivity issues that exist despite the use of the vector $S_\alpha(R)$, however, the computational costs have been prohibitively high. We conjecture that EA_{EPS} needs to grow its reference set just like the approximation-guided algorithm AGE [32] does.

4.2.2 Three-feature combinations. For three-feature combinations, the indicators examined are the hypervolume and the inverted generational distance. The results from optimizing these two indicators are compared with those from minimizing the discrepancy value. The statistics gathered from 30 repeated runs of each setting are included in Table 6. The three-feature combinations under examination in this paper are the same as in [23]. Table 6 summarizes the indicator values of the final populations after running the three algorithms on the three three-feature combinations. Both of the IGD values and HYP values of the final populations from EA_{IGD} and EA_{HYP} are better than those from EA_{DIS}. Although both algorithms do not perform very well in minimizing discrepancy for most three-feature combinations, EA_{HYP} is able to achieve a smaller average discrepancy value than EA_{DIS} in feature combination (f_1,f_3,f_4) and a comparable average value in feature combination (f_2,f_3,f_4). The minimum discrepancy values obtained by EA_{HYP} for the three different feature combinations are all smaller than the corresponding values from EA_{DIS}.

5 CONCLUSIONS

We have proposed a new approach for evolutionary diversity optimization. It bridges the areas of evolutionary diversity optimization and evolutionary multi-objective optimization and shows how techniques developed in evolutionary multi-objective optimization can be used to come up with diverse sets of solutions of high quality for a given single-objective problem. Our investigations demonstrated that well-established multi-objective performance indicators can be used to achieve a good diversity of sets of solutions according to a given set of features. The advantages of our approaches are (i) their simplicity and (ii) the quality of diversity achieved as measured by the respective indicators. The best performing approaches use HYP or IGD as indicators. We have shown that they achieve excellent results in terms of all indicators and often even outperform the discrepancy-based approach [23] when measuring quality in terms of discrepancy. This is surprising as they are not tailored towards this measure.

REFERENCES

[1] Bradley Alexander, James Kortman, and Aneta Neumann. 2017. Evolution of artistic image variants through feature based diversity optimisation. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 171–178.

[2] David Applegate, William Cook, Sanjeeb Dash, and André Rohe. 2002. Solution of a Min-Max Vehicle Routing Problem. *INFORMS Journal on Computing* 14, 2 (April 2002), 132–143.

[3] Rudolf Berghammer, Tobias Friedrich, and Frank Neumann. 2012. Convergence of set-based multi-objective optimization, indicators and deteriorative cycles. *Theor. Comput. Sci.* 456 (2012), 2–17.

[4] Karl Bringmann, Tobias Friedrich, Frank Neumann, and Markus Wagner. 2011. Approximation-Guided Evolutionary Multi-Objective Optimization. In *21nd International Joint Conference on Artificial Intelligence*. IJCAI/AAAI, Barcelona, Spain, 1198–1203.

[5] Shelvin Chand and Markus Wagner. 2015. Evolutionary many-objective optimization: A quick-start guide. *Surveys in Operations Res. and Management Science* 20, 2 (2015), 35 – 42.

[6] Raymond Chiong, Thomas Weise, and Zbigniew Michalewicz (Eds.). 2012. *Variants of Evolutionary Algorithms for Real-World Applications*. Springer.

[7] Gregory W. Corder and Dale I. Foreman. 2009. *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. Wiley.

[8] Kalyanmoy Deb. 2001. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Inc.

[9] Kalyanmoy Deb. 2012. *Optimization for Engineering Design - Algorithms and Examples, Second Edition*. PHI Learning Private Limited.

[10] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolutionary Computation* 6, 2 (2002), 182–197.

[11] Eelco den Heijer and AE Eiben. 2014. Investigating aesthetic measures for unsupervised evolutionary art. *Swarm and Evolutionary Computation* 16 (2014), 52–68.

[12] Wanru Gao, Samadhi Nallaperuma, and Frank Neumann. 2015. Feature-Based Diversity Optimization for Problem Instance Classification. *CoRR* abs/1510.08568 (2015). arXiv:1510.08568

[13] Wanru Gao, Samadhi Nallaperuma, and Frank Neumann. 2016. Feature-Based Diversity Optimization for Problem Instance Classification. In *Parallel Problem Solving from Nature, PPSN (LNCS)*. 869–879.

[14] John F Hughes, Andries Van Dam, James D Foley, Morgan McGuire, Steven K Feiner, David F Sklar, and Kurt Akeley. 2014. *Computer graphics: principles and practice*. Pearson Education.

[15] Ke Li, Kalyanmoy Deb, Qingfu Zhang, and Sam Kwong. 2015. An Evolutionary Many-Objective Optimization Algorithm Based on Dominance and Decomposition. *IEEE Trans. Evolutionary Computation* 19, 5 (2015), 694–716.

[16] Kresimir Matkovic, László Neumann, Attila Neumann, Thomas Psik, and Werner Purgathofer. 2005. Global Contrast Factor-a New Approach to Image Contrast. *Computational Aesthetics* 2005 (2005), 159–168.

[17] Olaf Mersmann, Bernd Bischl, Heike Trautmann, Markus Wagner, Jakob Bossek, and Frank Neumann. 2013. A novel feature-based approach to characterize algorithm performance for the traveling salesperson problem. *Annals of Mathematics and Artificial Intelligence* 69, 2 (2013), 151–182.

[18] Mario A. Muñoz, Laura Villanova, Davaatseren Baatar, and Kate Smith-Miles. 2018. Instance spaces for machine learning classification. *Machine Learning* 107, 1 (2018), 109–147.

[19] Samadhi Nallaperuma, Markus Wagner, and Frank Neumann. 2014. Parameter Prediction Based on Features of Evolved Instances for Ant Colony Optimization

[20] Samadhi Nallaperuma, Markus Wagner, and Frank Neumann. 2015. Analyzing the Effects of Instance Features and Algorithm Parameters for Max-Min Ant System and the Traveling Salesperson Problem. *Frontiers in Robotics and AI* 2 (2015), 18. https://doi.org/10.3389/frobt.2015.00018

[21] Samadhi Nallaperuma, Markus Wagner, Frank Neumann, Bernd Bischl, Olaf Mersmann, and Heike Trautmann. 2013. A Feature-based Comparison of Local Search and the Christofides Algorithm for the Travelling Salesperson Problem. In *Proceedings of the Twelfth Workshop on Foundations of Genetic Algorithms XII (FOGA XII '13)*. ACM, 147–160. https://doi.org/10.1145/2460239.2460253

[22] Aneta Neumann, Bradley Alexander, and Frank Neumann. 2017. Evolutionary Image Transition Using Random Walks. In *Computational Intelligence in Music, Sound, Art and Design - 6th International Conference, EvoMUSART (LNCS)*, Vol. 10198. 230–245.

[23] Aneta Neumann, Wanru Gao, Carola Doerr, Frank Neumann, and Markus Wagner. 2018. Discrepancy-based evolutionary diversity optimization. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 991–998.

[24] Aneta Neumann, Zygmunt L. Szpak, Wojciech Chojnacki, and Frank Neumann. 2017. Evolutionary image composition using feature covariance matrices. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 817–824.

[25] Mark Nixon and Alberto S. Aguado. 2008. *Feature Extraction & Image Processing, Second Edition* (2nd ed.). Academic Press.

[26] R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[27] Sebastian Risi, Sandy D. Vanderbleek, Charles E. Hughes, and Kenneth O. Stanley. 2009. How novelty search escapes the deceptive trap of learning to learn. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 153–160.

[28] Kenneth O. Stanley and Joel Lehman. 2015. *Why Greatness Cannot Be Planned - The Myth of the Objective*. Springer.

[29] Eric Thiémard. 2001. An Algorithm to Compute Bounds for the Star Discrepancy. *J. Complexity* 17 (2001), 850–880.

[30] Tamara Ulrich, Johannes Bader, and Eckart Zitzler. 2010. Integrating decision space diversity into hypervolume-based multiobjective search. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 455–462.

[31] Tamara Ulrich and Lothar Thiele. 2011. Maximizing population diversity in single-objective optimization. In *Genetic and Evolutionary Computation Conference, GECCO*. ACM, 641–648.

[32] Markus Wagner, Karl Bringmann, Tobias Friedrich, and Frank Neumann. 2015. Efficient optimization of many objectives by approximation-guided evolution. *European Journal of Operational Research* 243, 2 (2015), 465 – 479.

[33] Qingfu Zhang and Hui Li. 2007. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evolutionary Computation* 11, 6 (2007), 712–731.

[34] Eckart Zitzler and Simon Künzli. 2004. Indicator-Based Selection in Multiobjective Search. In *Parallel Problem Solving from Nature, PPSN (LNCS)*, Vol. 3242. Springer, 832–842.

[35] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. 2003. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evolutionary Computation* 7, 2 (2003), 117–132.