Many-Objective Optimization for Diverse Image Generation

Nathanaël Carraz Rakotonirina, Andry Rasoanaivo, Laurent Najman, Petr Kungurtsev, Jeremy Rapin, Fabien Teytaud, Baptiste Roziere, Olivier Teytaud, Markus Wagner, Pak-Kan Wong, et al.

▶ **To cite this version:**

Nathanaël Carraz Rakotonirina, Andry Rasoanaivo, Laurent Najman, Petr Kungurtsev, Jeremy Rapin, et al.. Many-Objective Optimization for Diverse Image Generation. 2021. hal-03425742

HAL Id: hal-03425742
https://hal.science/hal-03425742

Preprint submitted on 11 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Many-Objective Optimization for Diverse Image Generation

Nathanaël Carraz Rakotonirina[*a], Andry Rasoanaivo[†a], Laurent Najman[‡b], Petr Kungurtsev[§c], Jeremy Rapin[¶d], Fabien Teytaud[‖e], Baptiste Roziere[**d], Olivier Teytaud[††d], Markus Wagner[‡‡f], Pak-Kan Wong[§§g], and Vlad Hosu[¶¶h]

[a]Université d'Antananarivo
[b]LIGM, Univ Gustave Eiffel, CNRS, ESIEE Paris, F-77454 Marne-la-Vallée
[c]Enthought
[d]FAIR
[e]Univ. du Littoral
[f]University of Adelaide
[g]The Chinese University of Hong Kong
[h]University of Konstanz

November 11, 2021

Abstract

In image generation, where diversity is critical, people can express their preferences by choosing among several proposals. Thus, the image generation system can be refined to satisfy the user's needs. In this paper, we focus on multi-objective optimization as a tool for proposing diverse solutions. Multi-objective optimization is the area of research that deals with optimizing several objective functions simultaneously. In particular, it provides numerous solutions corresponding to trade-offs between different objective functions. The goal is to have enough diversity and quality to satisfy the user. However, in computer vision, the choice of objective functions is part of the problem: typically, we have several criteria, and their mixture approximates what we need. We propose a criterion for quantifying the performance in multi-objective optimization based on cross-validation: when optimizing $n-1$ of the n criteria, the Pareto front should include at least one good solution for the removed n^{th} criterion. After providing evidence for the validity and usefulness of the proposed criterion, we show that the diversity provided by multi-objective optimization is helpful in diverse image generation, namely super-resolution and inspirational generation.

[*]carraznathanael@live.fr
[†]r.andry.rasoanaivo@gmail.com
[‡]laurent.najman@esiee.fr
[§]pkungurtsev@enthought.com
[¶]jrapin@fb.com
[‖]teytaud@univ-littoral.fr
[**]broz@fb.com
[††]oteytaud@fb.com
[‡‡]markus.wagner@adelaide.edu.au
[§§]pkwong@link.cuhk.edu.hk
[¶¶]vlad.hosu@uni-konstanz.de

1 Introduction

Diversity in image generation is critical. Problems of insufficient representation of some classes can be mitigated by presenting several possibilities to the user, if there is enoough diversity. In the present paper, we propose multi-objective optimization as a tool for facilitating diverse image generation.

Multi-objective optimization (MOO [11, 17]) is the simultaneous optimization of several objective functions (OFs). For example, given a request, an internet search method typically ranks a list of possible answers using a combination of OFs. OF can include the relevance of the answer, price, geographical distance, reliability of the source, aesthetics, the technical quality of an image, etc. MOO can assist with this as it obtains a Pareto set of solutions, which is a set of solutions that are not dominated by other solutions (Eq. 1). Then, we can subsample that set and get a list of proposals [10] (Sec. 2.2). The user can then select a proposal. The goal of MOO is to help a user find the best solution according to their unknown OF. It can suggest either a huge but navigable set of solutions, or a small set that the user can exhaustively check: we focus on the latter.

While benchmarking single-objective optimization is already a complicated task, MOO raises additional issues: how to compare two MOO methods? The hypervolume (HV) indicator [2] (Eq. 2), which measures the dominated volume up to a reference point, is a simple criterion. However, the choice of the reference point is often arbitrary and can completely change the conclusion. In addition, this is not directly related to user satisfaction, and numerical comparisons based on HV might be biased in favor of methods using HV as a proxy.

The fact that we do not have an exact measure of user satisfaction is critical in MOO: if we had a preference function for guessing if the user prefers solution A or solution B, we would optimize that function. MOO precisely depends on having several proxies of the user satisfaction, e.g., cost, quality, maintainability, speed, etc., but not the exact measure. With an exact measure of satisfaction, we would not need MOO in the first place: a principled, practical comparison between MOO must take into account this approximate nature of the OFs.

We present several algorithms for MOO (Section 2.1), including specifically MOO methods, HV-based methods, and MOO variants of Differential Evolution (DE). We run all algorithms from Nevergrad [38] that can deal with MOO plus our new contributions. We also include several methods for subsampling Pareto sets. We propose a new method termed cross-validation of criteria, based on cross-validating objective functions, for benchmarking MOO (Section 3.1), specifically in the many-objective cases when the OFs are a redundant and partial approximation of the real user preference, so that cross-validating criteria makes sense.

We present computer vision benchmarks (Section 3.2), with, in some cases, a representation of images based on latent variables, and we report on extensive comparative experiments in computer vision (Section 4).

Finally, in Section 4.2, we visualize results for conditional image generation such as inspirational generative adversarial networks [4] (Section 4.2.1) and super-resolution (Section 4.2.2).

2 Multi-objective optimization

In this section, we start by presenting MOO and several algorithms that can deal with it (overview of algorithms in Table 1). We then present different methods for subsampling Pareto fronts (Section 2.2).

Name	Image repr.	k	$c_1,\dots,c_k = F \setminus \{u\}$	F'	M	Num. settings	Results
IQA assisted image generation							
Qualities	raw	3	$Q-\{q\}$	$K512$, Brisque	16	2	Figure 1
Qua. +Gan	latent	3	$Q-\{q\}$	$K512$, Brisque	16	2	Figure 2
Inspirational generation							
Similarity	raw	4	$S-\{s\}$	$s \in S$	25	$\|S\|=5$	Figure 1
Sim.+Gan	latent	4	$S-\{s\}$	$s \in S$	25	$\|S\|=5$	Figure 1
IQA assisted Inspirational generation							
Sim. + qua.	raw	7	$S+Q-\{q\}$	$K512$	16	1	Figure 2
Sim. + qua. +Gan	latent	7	$S+Q-\{q\}$	K512	16	1	Figure 3
Super-resolution							
Tarsier	latent	3	$K512$, $\|\|z\|\|$, D	Percep./PSNR	9	2	Figure 5, Tab. 2
Inspirational generation							
PytorchGanZoo	latent	3	$\|\|z\|\|$, D, S	Human	8	1	Figure 4

Name	MOO handling
(1+1)EA, PSO, F_{127}CMA (1+1)ES, NGOpt12	Convert to single using HV (Sec. 2.1.1)
NGOPt12/9, NGOpt12/16, NGOpt12/25	Convert to single using MSR (Sec. 2.1.2)
NSGA-II	Inherently MOO (Sec. 2.1.3)
DE	Inherently MOO (Sec. 2.1.4)

Table 1: **Top: different settings considered in the experimental comparisons.** Second col: in some cases, we use a latent representation of variables using a GAN (Section 3.3). Each row corresponds to several optimization settings (e.g., one for u =K512 and one for u =Brisque in the first row). k refers to the number of objective functions. u is the ground truth and M is the number of proposed solutions. IQA (image quality assessment) methods are presented in Section 3.2. The number of settings ($6^{th}col$) is the cardinal of F' (4^{th} row), i.e., the number of possible u (see Section 3.1). $||z||$ refers to the norm of the latent variable: this represents relevance to the domain as a large z means an image far from the center of the domain. D refers to the discriminator. S is a similarity defined in PytorchGanZoo (combining several measures). **Bottom: an overview of our optimization methods.** Each algorithm is equipped with one of the subsampling methods that extracts a Pareto front of limited size (Section 2.2). Unless stated otherwise, this is IGD.

2.1 Multi-objective optimization

Classical numerical optimization is the search for $x^* \in D$ minimizing some function $f : D \to \mathbb{R}$. Given N OFs $f_1, \dots f_N$, MOO is the (approximate) search for the Pareto front defined by $\{x^* \in D$ such that $\nexists x \in D, x \succ x^*\}$, where (in our minimization context) $x \succ x^*$ stands for Pareto-dominance

$$\forall 1 \leq i \leq N, f_i(x) \leq f_i(x^*) \wedge \exists j, f_j(x) < f_j(x^*). \tag{1}$$

Our study does not include all existing MOO. For example, we do not include approaches based on a representation by $\mu \times d$ variables of a Pareto approximation of cardinal μ in a domain of dimension d [53]. Also it does not consider preference elicitation or interactive MOO, for which the reader is referred to [3, 22, 56]. However, [17] distinguishes three categories of MOO, all of them represented here: (i) based on indicators, such as the HV (Section 2.1.1), (ii) decomposition by scalarization (Section 2.1.2), (iii) Pareto-based MOO, which use directly the Pareto rank and measures of contributions to the diversity (NSGA-II in Section 2.1.3 and a MOO variant of DE in Section 2.1.4). Importantly, due to repeated disappointing results when downloading/installing/adapting codes, we do not include any method that does not exist as a proper Pypi package.

Algorithm 1: The (1+1)-EA. x initialized at the center.

1. $x' \leftarrow x$
2. Mutation: $\forall 1 \leq i \leq d$: if $rand < \frac{1}{d}$, then $x'_i \leftarrow random \in \mathbb{R}$.
3. While $x' == x$ repeat the mutation above.
4. If x' better than x than $x \leftarrow x'$.
5. While we have time, go to 1.

2.1.1 Methods based on the HV

Given n points $x_1, \ldots, x_n$ and a reference point $y \in \mathbb{R}^N$, the HV indicator measures how they are distributed in the objective space. It is defined as

$$HV(x_1, \ldots, x_n, y) = \tag{2}$$
$$\mu(\{t \in R^n; \forall 1 \leq i \leq n, \forall 1 \leq j \leq N, f_j(x_i) \leq t_j \leq y_j\}),$$

with μ the Lebesgue measure. One can use the HV indicator for converting MOO into single-objective optimization. Importantly, the OF becomes dynamic: the objective value depends on previous points. With HV, we can use many existing algorithms. We refer to [38] for the full list of methods considered in the present paper; we mention Particle-Swarm Optimization (PSO [28]), Differential Evolution (DE [51]), Covariance Matrix Adaptation [23] (F127CMA refers to FCMA version 1.2.7), Diagonal-CMA (DCMA [42]), the $(1+1)$-ES with one-fifth rule [48], NGOpt12 the optimization wizard from Nevergrad automatically selecting a method depending on dimension/budget/type of variables. Inspired by [40], we also include algorithms from discrete optimization adapted to the continuous domains such as (1+1)EA. Describing all considered algorithms would be beyond the scope of the present paper: we only give an overview of the (1+1)EA in Alg. 1 because it is unusual in a continuous context.

A specificity of methods based on random mutations (such as the one in Alg. 1) is that they prefer flat stable basins rather than peaked small basins of attraction [13]. For this reason, they can be expected to generalize better from a proxy of objective function to the real objective function, which might be useful for MOO.

2.1.2 Multiple single-objective runs

When we want to find distinct trade-offs but do not want to bother playing with multiple OFs, we might use:

- MSRH (multiple single runs with handcrafted coefficients): use a human-defined trade-off, and optimize m times, assuming that the presence of multiple local minima will be enough for ensuring diversity so that the user will be happy with at least one of the results.
- MSR (multiple single runs): optimize, m times, a randomly chosen linear combination of weights. With $NGOpt12$ being the default optimizer in Nevergrad, we consider $NGOpt12_m$ the method running m times a single objective optimization with each of the original OFs weighted by a randomly uniformly drawn coefficient in $[0, 1]$.

2.1.3 Methods based on NSGA-II

NSGA-II [14] is a well-known generic method for transforming an optimization method based on selection and/or crossover into a MOO method. NSGA-II uses the crowding distance as a comparison tool: it prefers non-dominated solutions, but (if needed) it will also prefer solutions that occupy the largest cuboids in the objective space, where the size of the solution's cuboid is defined with respect to its neighbouring solutions.

2.1.4 Variants of differential evolution adapted to MOO

We also use a variant of differential evolution specifically dedicated to MOO. Our DE is based on the differential evolution code in [38], i.e., a rand-to-best variant, modified using ideas from [41] and [1]. DE is adapted to MOO as follows:

- For comparing a point with its parent: each time DE requests such a comparison, DE randomly chooses one of the k objective functions and returns True if the child outperforms its parent for this objective function.
- Selection of the best: each time we need "the" best point in DE, our DE randomly selects one of the Pareto optimal solutions in the population.

Nevergrad contains an optimization wizard which automatically chooses an optimization method. It turns out that when it detects that the problem is MOO (i.e., when it receives some objective values), then it switches to our implementation of DE. Therefore, it randomly samples the domain before switching to DE. This is termed DE+ in our results, and performs close to DE.

2.2 Approximating Pareto-fronts

To extract $1 \leq m \leq n$ points from an approximate Pareto set $\{x_1, \dots, x_n\}$, a range of approaches can be used:

- Random subset: just pick up m of the x_i, uniformly at random and without replacement.
- HV: pick up $\{x_{j_1}, \dots, x_{j_m}\}$ such that their Hypervolume C_h is maximal.
- Loss-covering, also known as IGD (inverted generational distance, [45]): pick up $\{x_{j_1}, \dots, x_{j_m}\}$ such that $C_l = \sum_{i=1}^{n} \inf_{j \leq m} ||F(x_i) - F(x_{i_j})||^2$ is minimal, where $F(x) = (f_1(x), \dots, f_N(x))$.
- Domain-covering: pick up $\{x_{j_1}, \dots, x_{j_m}\}$ such that $C_d = \sum_{i=1}^{n} \inf_{j \leq m} ||x_i - x_{i_j}||^2$ is minimal.
- Additive epsilon approximation (EPS, [35]): pick up $\{x_{j_1}, \dots, x_{j_m}\}$ such that $C_e = \max_{i=1}^{n} \inf_{j \leq m} ||F(x_i) - F(x_{i_j})||_\infty$ is minimal, where $F(x) = (f_1(x), \dots, f_N(x))$.

Finding optimal subsets for some of those OFs is known to be NP-hard [9]. While subset selection is an important research area, it is outside of the scope of this article. We therefore resort to a simple, unbiased approach: we randomly draw 30 subsets and pick the best for the chosen criterion.

3 Experimental setup: a new criterion and image generation

In the present section, we present tools necessary for the rest of the paper. Section 3.1 presents a new criterion for comparing MOO. Section 3.2 presents our computer vision benchmarks: some of those benchmarks use a latent representation of images, presented in Section 3.3.

3.1 Methodology for comparing MOO methods: cross-validating objective functions

There are two main families of MOO. First, sometimes, MOO is interactive [56]: the user provides feedback, possibly through preference elicitation [8], during the MOO run. Second, purely offline MOO: the algorithm is given some OFs $c_1, \dots, c_k$, where for all i, $C_i : X \to \mathbb{R}$ and outputs a finite set $\{p_1, \dots, p_M\} \subset X$. Then the user selects $p_{i^*} \in \{p_1, \dots, p_M\}$. We postulate that there exists a u which quantifies the user preference: $\forall i \in \{1, \dots, M\}$, $u(p_{i^*}) \leq u(p_i)$. The function u is related to the $c_1, \dots, c_k$ but is not one of them. We focus on this second scenario and propose a method for benchmarking performance in such a case. Many methods not directly connected to user satisfaction have been proposed for quantifying the performance of MOO [6]: our criterion is centered on the satisfaction of an unknown preference, related but not equal to the objective functions. More precisely, we will use an auxiliary objective function $u : X \to \mathbb{R}$ as a ground truth, and use

it as follows. **Cross-validated criterion (CVC (F,F')) for MOO:** given an integer M, a set F of objective functions and a subset $F' \subset F$ of possible targets, we will consider, for each target $u \in F'$, a scenario $scenario_{u,F\setminus\{u\}}$ as follows.

Given F a finite set of OFs, u one of them, we consider $\{c_1,\dots,c_k\}$ the set F after removing u. In $scenario_{u,(c_1,\dots,c_k)}$, we multi-optimize $(c_1,\dots,c_k)$. The output of the algorithm is then a finite approximation $(p_1,\dots,p_M)$ of the Pareto set. Our loss is $\inf_{1\leq i\leq M} u(p_i) - u(p^*)$, with p^* the best solution, for u, met in all our runs in $scenario_{u,(c_1,\dots,c_k)}$.

Typically, F' contains the OFs in F to be used as ground truth. For example, Koncept512 [24] is the best approximation of opinion scores provided by humans in terms of quality assessment, so it is in F', whereas pure Blurriness is not. Table 1 presents several such contexts used in our experiments.

3.2 Diverse computer vision

Cheng et al. [12] mention how artificial problems studied in the MOO literature are easier than real-world ones, which entails misleading benchmark results. We consider computer vision tasks combining some of the following categories of objective functions: quality assessment (often combining several scores), discriminator score, distance to the target, user feedback, and distance to the domain of latent variables. We consider the following similarity measures: $S = \{$distance between histograms, L1-distance, L2-distance, Lpips-AlexNet [29, 59], Lpips-Vgg [49]$\}$. We consider the following quality measures: $Q = \{$Blurriness, Brisque [33], Koncept512 [24]$\}$. To compare MOO algorithms, we use either the HV or our proposed cross-validated criterion (Section 3.1). We consider many settings, all for computer vision, as described in Table 1:

- Image generation assisted by image quality assessment, as in e.g. [44]. The idea here is to improve the quality of the images generated by a generative adversarial network (GAN). The image quality assessment (IQA) in such tools is sometimes performed by deep networks [24] or by interaction with a human (HEVOL in [40]): this automatization of quality assessment removes tedious manual search over thousands of generated images [50].
- Inspirational generation as in e.g. [40]. Inspirational generation consists of generating an image, in a given domain typically represented by a dataset, close to another given image, by searching the latent space. The OF of similarity to a given image is the difference with a classical GAN. For example, consider toonme.com, ranked #1 on the app store at the time of the present writing: this app takes as input a face image, and generates multiple images as an output (i) similar to the input (ii) in a cartoon domain, for "toonifying" that face. This follows the tradition of applying GANs for design [47, 61]. Compared to [20, 36, 39], this method searches a latent space and provides an output in the training domain. Compared to [31], this does not modify the training procedure.
- Inspirational generation assisted by image quality assessment. For adding constraints on the generated image of a GAN, style transfer constrains the training [20, 36]. Other approaches analyze the latent space [40]: a z matching some constraints or optimizing some OF is used, in lieu of a randomly generated z. This means that we optimize simultaneously both the quality and the similarity to some example. Such an approach is directly useful for artists or designers [55], but also for inpainting, facial composites, anonymization or photo edition [7, 19, 52, 57].

We have cases in which we work on raw images and cases in which we work on latent variables as detailed in Section 3.3.

3.3 Image representation by GAN latent variables

Generative models, and in particular adversarial ones [21], are becoming prevalent in computer vision as they enhance artistic creation [16, 60, 36], inspire designers [47, 61], or prove useful in semi-supervised learning [15, 18, 34]. GANs typically generate random images. More precisely, a random latent vector z is

generated, and the generator outputs $G(z)$. Much like in some previously mentioned works, we use z as a smaller and more structured representation of an image.

4 Experimental results

For most experiments, we rely on the open-source project Nevergrad[1]. We did some small modifications to PytorchGanZoo[2] and Tarsier[3]. In numerical minimization, the loss means the objective value of the best candidate, minus the minimum possible objective value. In all experimental results, we use plots as provided by Nevergrad [38], namely average (over settings) median (over replicas) loss. Before averaging, the loss is linearly normalized to $[0,1]$. The definition of loss (a.k.a. simple regret) refers to an optimal value: this is replaced by the best-known value when the optimal value is unknown. Light-colored lines refer to $\pm$ standard deviation. For our new CVC (Section 3.1), we plot the average loss for this CVC. Other plots are based on the HV and we plot the average loss for the HV. The budget refers to the number of candidate points evaluated by the OFs. In all plots, the lower, the better. The quality assessment tools can fail: Figure 4 shows that PytorchGanZoo can produce aberrant data: using MOO precisely helps for mitigating such issues by considering several trade-offs. As we use the median, missing data means that at least 50% of replicas failed.

4.1 Computer vision benchmarks: HV and CVC

Figures 1-3 present experimental results. Roughly speaking, the MOO variant of DE we propose, based on [1] and [41], performs well in many cases. However, in the case of CDC for Many-OO (Figure 3, middle), MSR is sometimes better. These facts are consistent with our observations on PytorchGanZoo (Section 4.2.1) and Tarsier (Section 4.2.2). In some difficult ill-defined cases such as in Figure 1, HV with (1+1)EA is best: random mutations are known for focusing on wide, flat, stable basins. This confirms [13] and, closer to our field, a remark in [40] and in Section 2.1.1: with $(1+1)$-EA (also known as Discrete-$(1+1)$) we get solutions which, though not better numerically, are preferred by humans. For subsampling methods, IGD performs best overall (Figure 3, right).

4.2 Examples of applications: diverse inspirational generation and diverse super-resolution

4.2.1 Inspirational generation

Inspirational generation consists in the following: given a target image t, generating an image $i(t)$, which is in a given domain D while having similarities with t (e.g. we look for a human face $i(t)$ with some similarities with Casimir (t) the gentle orange dinosaur). We use the code from [40]. A popular application of diverse image generation inspired by a target image is toonme.com (based on [26]). [40] uses a weighted sum of 3 OFs, namely image similarity to t, relevance to D (quantified by the norm of the latent injection), and realism according to the GAN discriminator. We test:

- multiobjective methods equipped with the various Pareto sampling tools in Section 2.2;
- MSRH, as it turns out that the problem is so multimodal that there is quite a lot of diversity.

We select the two methods that seem most promising, namely MSRH and DE. Then, we compare them through a double-blind human study. Results of the human study show that the latter (and simpler) method performs best: $68.57\% \pm 7.8\%$ (35 generated groups of 8 images by MSRH vs 8 images generated by DE

[1] https://github.com/facebookresearch/nevergrad
[2] https://github.com/facebookresearch/pytorch_GAN_zoo
[3] https://github.com/ncarraz/ESRGANplus

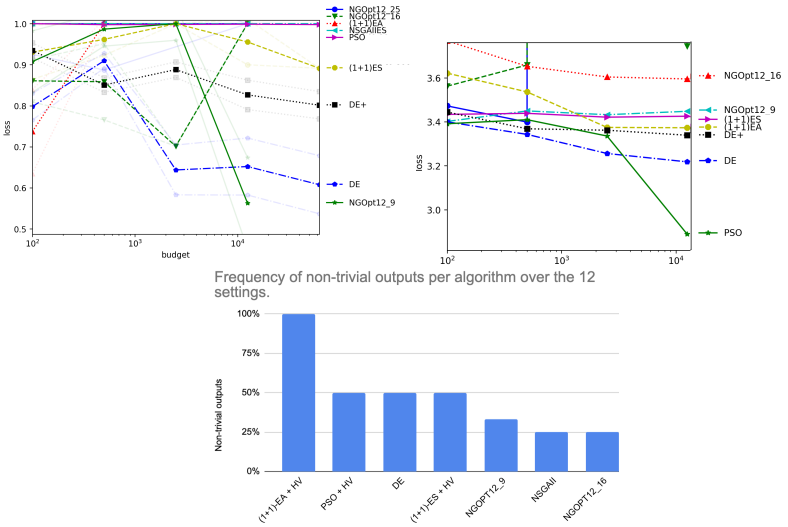


Figure 1: (Bigger version in the Supplementary Material) Various experiments: the $(1+1)$-EA is best (as in [40]) when there is a risk of optimization-based artifacts (quality optimization), and MOO DE is best overall in other cases. **Left, similarity optimization on raw images:** MOO-optimizing the five similarity measures. We compare algorithms using the HV. The best methods, namely NGOpt12/9 (i.e. a MSR) and variants of DE based on MOO-specific operators as in Section 2.1.4, are actually not based on HV. **Middle, similarity optimization through PGAN** [25]: CVC. Average similarity (renormalized as detailed in Section 4) for one of the similarity measures when MOO-optimizing the four other similarity measures (see Section 3.2) with a latent representation. The MOO DE from Section 2.1.4 performs among the best for many values of the budget. **Right, quality optimization: CVC.** We have 12 contexts made of two settings and 6 budget values: we compute the best K512 (resp. Brisque in the second setting) score among the 16 images obtained by MOO of Blurriness and Brisque (resp. K512 in the second setting). We present the frequency at which non-trivial (finite) values were obtained over those 12 cases. This case is challenging, as the quality OFs are quite different from each other (Koncept512, based on a neural net for IQA; Brisque; and pure Blurriness) - for example, in many cases, we get a failure as Brisque value for the 16 output images of MOO optimizing Koncept512 and Blurriness, hence missing results. The only method which provided a non-trivial median result in all cases, consistently with results in [40], is (1+1)-EA: this algorithm, originally from the discrete optimization community, prefers optima with stable flat basins (Section 2.1.1). Fischer's exact test: p-value $< 1.4\%$ for (1+1)-EA vs any other.

with IGD: the best image is one of the MSRH in 24 cases - p-value 0.05 for the exact Clopper-Pearson test). Figure 4 shows results for this repeated single objective optimization.

4.2.2 Super-resolution

Super-resolution is known to be an ill-posed problem, because a low-resolution image can correspond to multiple high-resolution versions. We apply MOO to image super-resolution in order to generate several super-resolved images instead of just one. Tarsier [43] is a GAN-based and perceptual-driven [30, 37] super-resolution model that uses noise injections in its architecture. At inference time, the injected noise is optimized according to two OFs along with an l_2 penalization: image quality score based on the IQA Koncept512 [24] and realism score based on the discriminator. We therefore have three objective functions. According to our previous results with CVC, our MOO DE might perform better than DCMA (Diagonal-CMA): We run the two methods to experiment on images of the Set14 [58] and Set5 [5] datasets which are benchmark datasets for super-resolution. As presented in Figure 5, the images obtained by DE are more diverse compared to those obtained using DCMA, and even more compared to images obtained by MSRH which are essentially identical: here MOO, using classical Pareto-dominance, does provide diversity. This is further confirmed quantitatively by the perceptual metric [59] of the Set5/Set14 images presented in Tabs 2

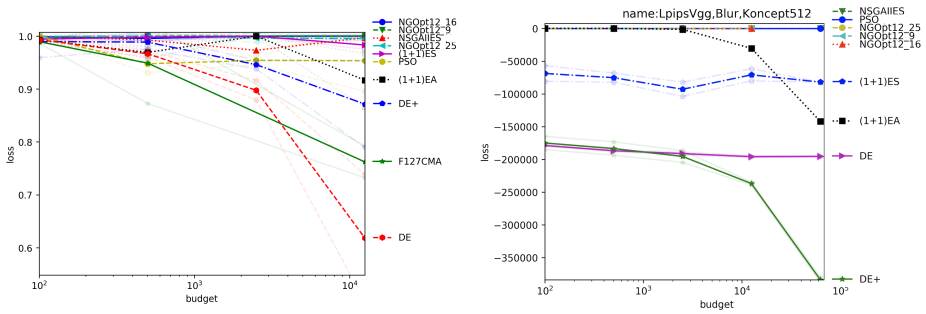


Figure 2: Other experiments, also good performance for MOO DE. **Left: quality optimization through PGAN, HV evaluation.** Aggregation (as detailed in Section 4) of two settings, K512 (resp. Brisque) score of an image obtained by MOO of Blurriness and Brisque (resp. K512), with latent variables as a representation. The y-axis is the (normalized and averaged, as explained in Section 4) HV: once more, DE performs best, though it is not based on the HV. **Right: similarity and quality optimization: CVC.** Best Lpips-Alex obtained by various algorithms in budget up to 5e4 when MOO-optimizing the similarity measures and the following quality measures: Lpips-Vgg, Bluriness, Koncept512. For clarity and ease of reading, only the best performing methods are presented on each plot.

	DCMA				DE			
	PSNR		Perceptual		PSNR		Perceptual	
	max (best)	std	min (best)	std	max (best)	std	min (best)	std
baby	28.5013	0.0584	0.1258	0.0012	28.5942	0.0284	0.1170	0.0016
bird	28.4378	0.1615	0.0487	0.0019	28.5327	0.0920	0.0464	0.0013
butterfly	23.1786	0.0208	0.0540	0.0003	23.1997	0.065	0.0534	0.0007
head	26.8428	0.0098	0.1190	0.0006	26.7854	0.0496	0.1150	0.0018
woman	25.4202	0.0231	0.0818	0.0003	25.4 513	0.1113	0.0814	0.0019

Table 2: **Peak signal-to-noise ratio (PSNR) and perceptual score of images in Set5.** We compare DE (MOO variant) and DCMA (based on HV) and check if the good performance of DE is confirmed here. **PSNR is to be maximized** and we see that in all cases but one (head) the best value for DE's outputs is better than the best value for DCMA's outputs. **Perceptual is to be minimized**: the best value (min) for DE's outputs is always better than the best value for DCMA's outputs. Consistent with observations in the rest of this paper, we get a better best value among the nine outputs of DE than among the nine outputs of DCMA with HV. Neither PSNR nor Perceptual were optimized for obtaining those images: this is an application of CVC, i.e., we check on other criteria than those that were optimized. Human assessment (Figure 5) confirms DE's superior performance. Conclusions: the MOO variant of DE does a good job here, and the variance of results over the selected Pareto front is significant; therefore, MOO does work for bringing diversity in super-resolution.

and 3. The peak signal-to-noise ratio (PSNR) is also provided for reference. This approach, which encourages diversity, could improve fairness [27, 32, 46] in generative models.

5 Conclusion

We advocate multi-objective optimization (MOO) as a means for generating diverse solutions to computer vision problems. The approach improves user satisfaction and fairness. Multiple solutions stand a better chance of satisfying the more demanding users. After performing experiments in Nevergrad, we validate our results in Super-Resolution (Tabs 2-3 and Figure 5) and Inspirational generation (Figure 4). **Depending on the case, we advocate one of two different MOO strategies: pareto fronts and MSR.** Many MOO problems arising in computer vision use ill-defined criteria, which do not exactly match human preference. In such cases, defining solutions based on Pareto fronts is questionable. For instance, if a dominates b it does not always follow that a is better than b. Typically, image quality, as a proxy for user satisfaction, might favor solutions that represent the most frequent categories, i.e., objects/people, leading to

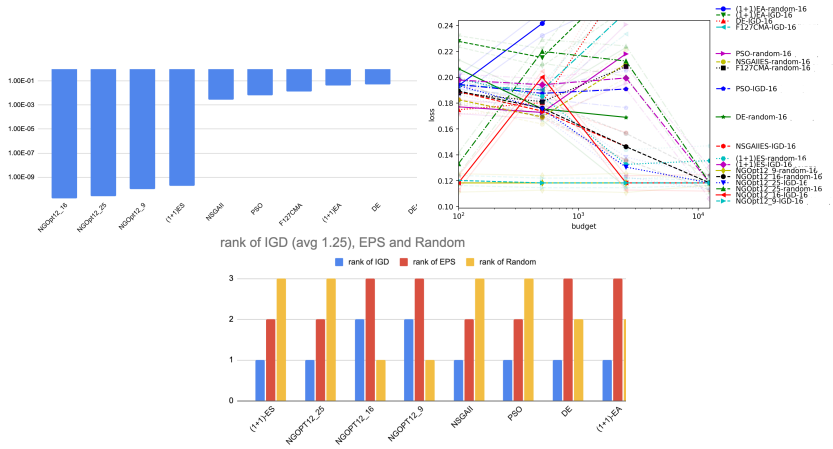


Figure 3: (Bigger version in the Supplementary Material) Settings close to inspirational generation (hence diversity is critical): MSR performing best, consistent with later experiments. **Many-objective similarity and quality optimization, through PGAN.** Images are represented by latent variables of a GAN. **Left, median HV** (log HV divided by max) obtained for each algorithm with budget 1e4: DE variants perform best, as in our artificial experiments. **Middle, comparison based on the CVC methodology:** best K512 score over images obtained by various MOO treatments of several similarity metrics and Brisque and Blurriness. We see that only MSR (and, interestingly, all variants of MSR) provides stable results in this inspirational generation context. This is consistent with previous results: whereas HV incorrectly predicts better results for DE. **Right: impact of the subsampling method** extracting M non-dominated points. For each algorithm, we compute the CVC loss, normalized and averaged. For each algorithm, we show the rank of the best result obtained by IGD (resp. EPS and Random) as subsampling methods. IGD significantly outperforms EPS (Wilcoxon p-value < 0.05).

an unfair representation of the others. This makes the MSR strategy (and MSRH when a tuned combination of criteria is available and multiple local optima are present) particularly competitive as they aim to optimize locally. This subtle point is critical for increasing diversity in inspirational generation, as shown by Figs. 3 and 4. However, MSR does not always produce the best results: for conditional GANs such as those used for super-resolution, MSR leads to nearly identical images, and in this context "real" "Pareto-style" MOO helps (Figs. 5 and Tab. 2). **Next, we propose a new principled methodology, termed CVC, for comparing MOO methods by cross-validation of objective functions.** In Section 3.1 we advocate this methodology for the many-objective cases. Compared to the HV method, CVC is not, and by design can not be based on an indicator used by some algorithms. Therefore, it facilitates an unbiased comparison between methods. For the many-objective cases, with partially redundant objectives designed as proxies for user preference, CVC helps for investigating the sufficiency of a set of objective functions. This is particularly relevant in computer vision, where we often use multiple similarity measures, which are but rough approximations of the criteria implemented by the human visual system. Results obtained with CVC are consistent with human inspection and diversity measures: Sections 4.2.1 and 4.2.2 show that when a method is "good" according to the CVC criterion, it is typically also preferred by users. Results using CVC in Tabs 2-3 are consistent with human inspection in Figure 5 and Figs. 3 and 4 are also in accordance. Note that human studies in scientific papers are a form of CVC: we optimize for some OF and validate with another OF, namely the user preference. **We open source a large family of benchmarks for MOO in computer vision.** They are merged in a maintained platform and can be run as one-liners. **Limitations.** We compare numerous methods in our comparative results (Section 4.1), reproducible in one-line in Nevergrad. However, the field is wide and, as discussed in [11], the best algorithms are very problem-dependent so we could include even more algorithms. Our open sourced MOO DE is heavily inspired by [41, 1] (Section 2.1.4). Our results in super-resolution do not include datasets focused on human faces yet, even though diversity is particu-

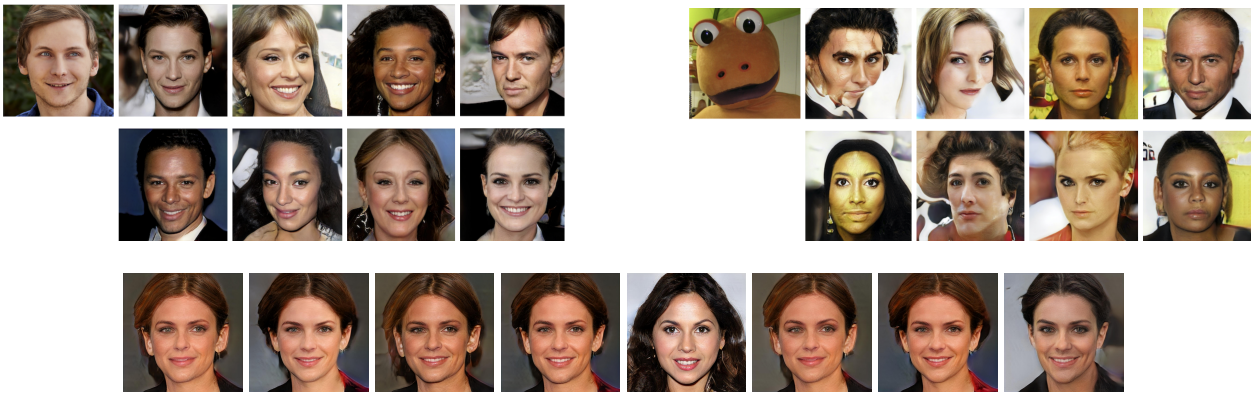
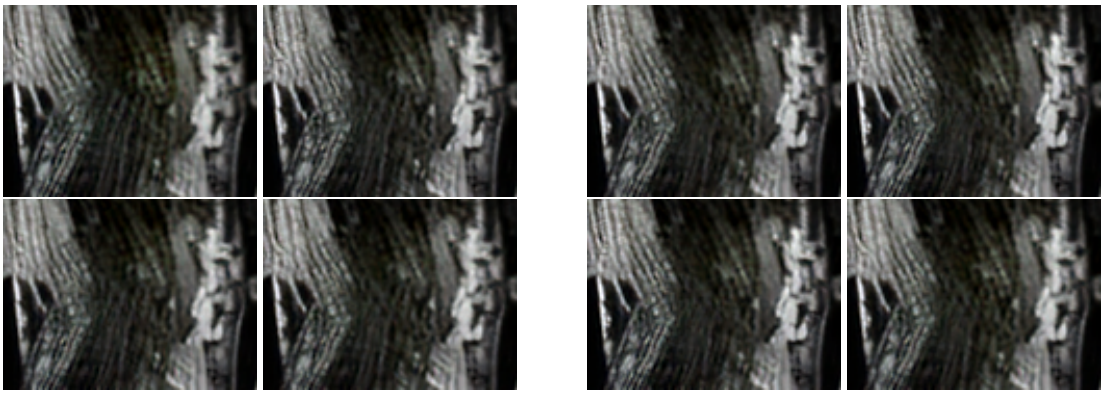


Figure 4: (Extended version in the Supplementary Material) **MSRH for inspirational generation (projection onto the celebrities model): two examples in which (1) MOO optimization provides diversity (2) MOO by MSRH outperforms MOO by criteria using Pareto-dominance (3 versions).** In the context of inspirational generation (generating an image similar to a target), it turned out that the local optima of the original trade-off optimization from [40], obtained by repeated optimization runs (MSRH method), offer more diversity and quality than other multiobjective methods (see text: OF are valid only locally, leading to a diversity loss when applying Pareto-dominance globally, hence the success of MSR variants). MSR is already the best for CVC in Figure 3. In these two cases above, the target image is the top left one. We present two hard cases, on which the original PytorchGanZoo code (without MOO) frequently fails, and PytorchGanZoo with Pareto-dominance generates very little diversity. **Top left:** the original inspirational GAN tends to generate images of women whereas the target is male; in contrast to this, over the eight obtained images, our MOO code generates male faces. **Top right:** in this difficult case, we look for a face with similarities with Casimir [54], the gentle orange dinosaur. Two of the faces have the orange color and the big dark-circled white eyes. Three of the eight generated faces are complete failures, but with the diversity obtained by MOO the user can select the best of the eight generated images. **Bottom, unsuccessful inspirational generation using Pareto dominance, compared to top left:** we get very little diversity (all female). Conclusions: there is an intrinsic variance in inspirational generation so that MSRH does work quite well: using MSRH rather than applying Pareto-dominance means using objective functions only locally and avoids discarding dominated parts of the domain.

larly important in that case for fairness reasons. Our proposed CVC criterion is meaningful only for many redundant misspecified OFs.

	DCMA				DE			
	PSNR		Perceptual		PSNR		Perceptual	
	max (best)	std	min (best)	std	max (best)	std	min (best)	std
baboon	**18.5173**	0.0302	0.2103	0.0018	18.4682	0.0361	**0.2072**	0.0029
bridge	**20.8810**	0.0540	0.3957	0.0026	20.8514	0.0194	**0.3942**	0.0034
comic	**19.7938**	0.0613	**0.1601**	0.0022	19.7166	0.0311	0.1622	0.0018
flowers	**23.4038**	0.0626	**0.1205**	0.0011	23.3265	0.0780	0.1212	0.0019
lenna	27.5146	0.1373	**0.0907**	0.0044	**27.6251**	0.2288	0.0933	0.0050
monarch	**30.0587**	0.0548	**0.0513**	0.0010	30.0167	0.0676	0.0517	0.0004
ppt3	**24.4132**	0.0722	0.0505	0.0006	24.3679	0.0369	**0.0504**	0.0004
barbara	**23.1106**	0.0607	**0.2699**	0.0006	23.0858	0.0710	0.2714	0.0006
coastguard	21.2996	0.0539	**0.2329**	0.0046	**21.3308**	0.0388	0.2364	0.0017
face	25.6842	0.0236	**0.1089**	0.0032	**25.6920**	0.0371	0.1119	0.0034
foreman	**27.8233**	0.0509	0.0805	0.0008	27.7125	0.0671	**0.0792**	0.0020
man	24.0318	0.0735	0.2136	0.0004	**28.3252**	0.0375	**0.0657**	0.0010
pepper	**28.3480**	0.0805	**0.0636**	0.0018	27.6251	0.2288	0.0933	0.0050
zebra	**24.4611**	0.1376	**0.1072**	0.0019	24.3468	0.0696	0.1088	0.0016

Table 3: **Peak signal-to-noise ratio (PSNR) and perceptual score of images in Set14.** We get a significant diversity, whereas running MSRH leads to several times the exact same image. Here, overall and contrarily to Tab. 2, Diagonal-CMA performs slightly better than DE.



(a) DE (b) DCMA

Figure 5: **Comparison between DCMA (diagonal CMA optimizing HV) and MOO DE.** Zooms on high-resolution images obtained by MOO-Tarsier using DE (a) and DCMA (b) on one image in Set14. In many cases, as in the present example, the images generated by DE are more diverse. For DE, the first image is blurrier and darker than the second one. The textures are all different for the four DE images.

References

[1] H. A. Abbass, R. Sarker, and C. Newton. PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems. In Congress on Evolutionary Computation (CEC), volume 2, pages 971–978, 2001.

[2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications. Theoretical Computer Science, 425:75 – 103, 2012. Theoretical Foundations of Evolutionary Computation.

[3] V. Belton and T. Stewart. Multiple criteria decision analysis: an integrated approach. Springer, 2002.

[4] D. Berthelot, T. Schumm, and L. Metz. BEGAN: boundary equilibrium generative adversarial networks. CoRR, abs/1703.10717, 2017.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In British Machine Vision Conference, pages 135.1–135.10. BMVA Press, 2012.

[6] D. Bigaud, X. Yu, Y. Lu, and X. Yu. Evaluating multiobjective evolutionary algorithms using MCDM methods. Mathematical Problems in Engineering, 2018:9751783, 2018.

[7] P. Bontrager, W. Lin, J. Togelius, and S. Risi. Deep interactive evolution. Computational Intelligence in Music, Sound, Art and Design - 7th International Conference, EvoMUSART, 2018.

[8] J. Branke, K. Deb, K. Miettinen, and R. Słowiński, editors. Multiobjective Optimization: Interactive and Evolutionary Approaches. Springer, 2008.

[9] K. Bringmann, S. Cabello, and M. T. M. Emmerich. Maximum Volume Subset Selection for Anchored Boxes. In 33rd International Symposium on Computational Geometry (SoCG), pages 22:1–22:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

[10] D. Carmel, E. Haramaty, A. Lazerson, and L. Lewin-Eytan. Multi-objective ranking optimization for product search using stochastic label aggregation. In Web Conference (WWW), page 373–383. ACM, 2020.

[11] S. Chand and M. Wagner. Evolutionary many-objective optimization: A quick-start guide. Surveys in Operations Research and Management Science, 20(2):35 – 42, 2015.

[12] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff. Test problems for large-scale multiobjective and many-objective optimization. IEEE Transactions on Cybernetics, 12:4108–4121, 2017.

[13] K. A. De Jong. Genetic algorithms are not function optimizers. In Foundations of Genetic Algorithms, volume 2, pages 5 – 17. Elsevier, 1993.

[14] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation, 6(2):182–197, 2002.

[15] J. Donahue, P. Krähenbühl, and T. Darrell. Adversarial feature learning. International Conference on Learning Representations (ICLR), 2017.

[16] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone. Creative adversarial networks. International Conference on Computational Creativity (ICCC), 2017.

[17] M. T. Emmerich and A. H. Deutz. A tutorial on multiobjective optimization: Fundamentals and evolutionary methods. Natural Computing, 17(3):585–609, Sept. 2018.

[18] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321:321–331, 2018.

[19] O. Gafni, L. Wolf, and Y. Taigman. Live face de-identification in video. In IEEE/CVF International Conference on Computer Vision (ICCV), pages 9378–9387, 2019.

[20] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. Conference and Workshop on Neural Information Processing Systems (NeurIPS), 2014.

[22] S. Greco, J. Figueira, and M. Ehrgott. Multiple criteria decision analysis, volume 37. Springer, 2016.

[23] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation, 11(1), 2003.

[24] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing, 29:4041–4056, 2020.

[25] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. International Conference on Learning Representations (ICLR), 2018.

[26] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of StyleGAN. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[27] P. J. Kenfack, D. D. Arapovy, R. Hussain, S. Kazmi, and A. M. Khan. On the fairness of generative adversarial networks (gans). arXiv preprint arXiv:2103.00950, 2021.

[28] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In IEEE International Conference on Neural Networks, pages 1942–1948, 1995.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.

[30] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4681–4690, 2017.

[31] Q. Mao, H. Lee, H. Tseng, S. Ma, and M. Yang. Mode seeking generative adversarial networks for diverse image synthesis. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1429–1437. IEEE, June 2019.

[32] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635, 2019.

[33] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing, 21(12):4695–4708, 2012.

[34] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen. Medical image synthesis with context-aware generative adversarial networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017.

[35] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In 41st Annual Symposium on Foundations of Computer Science, pages 86–92, 2000.

[36] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu. Semantic image synthesis with spatially-adaptive normalization. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 2337–2346, 2019.

[37] N. C. Rakotonirina and A. Rasoanaivo. Esrgan+ : Further improving enhanced super-resolution generative adversarial network. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3637–3641. IEEE, 2020.

[38] J. Rapin and O. Teytaud. Nevergrad - A gradient-free optimization platform. https://GitHub.com/FacebookResearch/Nevergrad, 2018.

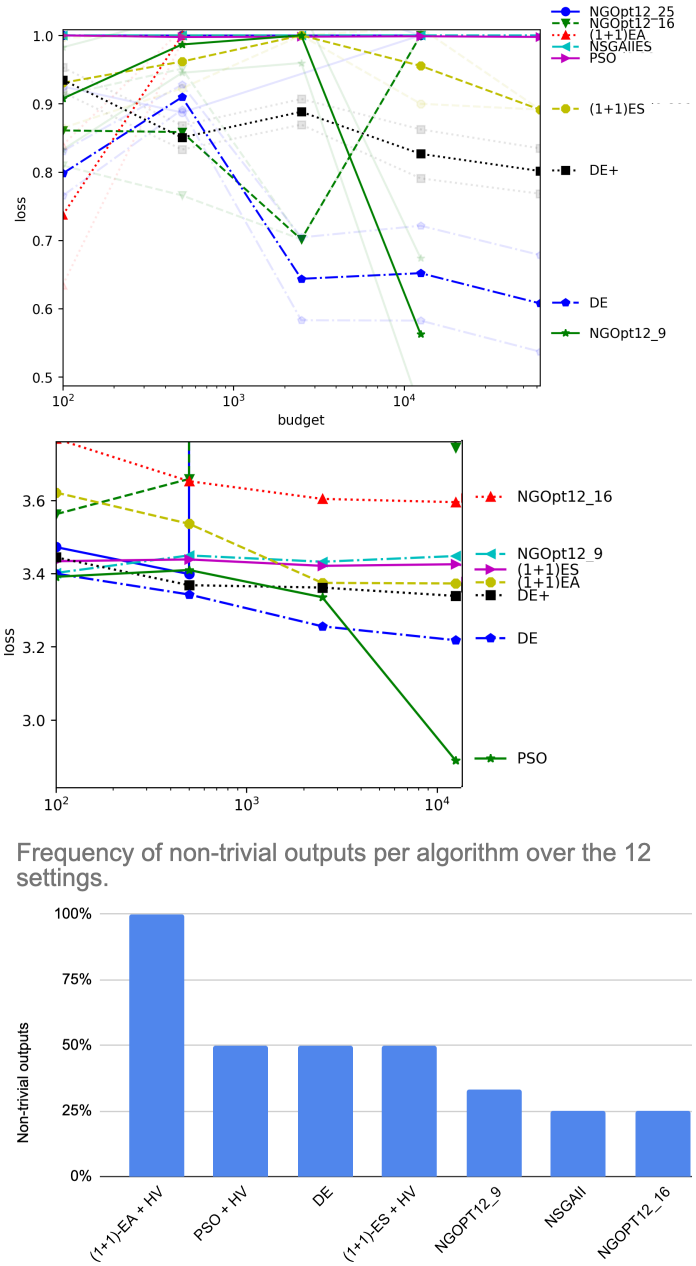
[39] S. E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee. Learning what and where to draw. In Conference and Workshop on Neural Information Processing Systems (NeurIPS). 2016.

[40] M. Riviere, O. Teytaud, J. Rapin, Y. LeCun, and C. Couprie. Inspirational adversarial image generation. arXiv preprint 1906.11661, 2019.
[41] T. Robič and B. Filipič. DEMO: Differential evolution for multiobjective optimization. In Evolutionary Multi-Criterion Optimization, pages 520–533. Springer, 2005.
[42] R. Ros and N. Hansen. A simple modification in cma-es achieving linear time and space complexity. In Parallel Problem Solving from Nature (PPSN), pages 296–305. Springer, 2008.
[43] B. Roziere, N. C. Rakotonirina, V. Hosu, A. Rasoanaivo, H. Lin, C. Couprie, and O. Teytaud. Tarsier: Evolving noise injection in super-resolution gans. In 25th International Conference on Pattern Recognition (ICPR), pages 7028–7035, 2021.
[44] B. Roziere, F. Teytaud, V. Hosu, H. Lin, J. Rapin, M. Zameshina, and O. Teytaud. Evolgan: Evolutionary generative adversarial networks. In Asian Conference on Computer Vision (ACCV), November 2020.
[45] H. Sato, H. E. Aguirre, and K. Tanaka. Local dominance using polar coordinates to enhance multiobjective evolutionary algorithms. In Congress on Evolutionary Computation (CEC), volume 1, pages 188–195 Vol.1, 2004.
[46] P. Sattigeri, S. C. Hoffman, V. Chenthamarakshan, and K. R. Varshney. Fairness gan: Generating datasets with fairness properties using a generative adversarial network. IBM Journal of Research and Development, 63(4/5):3–1, 2019.
[47] O. Sbai, M. Elhoseiny, A. Bordes, Y. LeCun, and C. Couprie. DesIGN: Design Inspiration from Generative Networks. ECCV workshop on Fashion, Art and Design, 2018.
[48] M. Schumer and K. Steiglitz. Adaptive step size random search. IEEE Transactions on Automatic Control, 13(2):270–276, 1968.
[49] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[50] E. L. Spratt. Creation, curation, and classification: Mario klingemann and emily l. spratt in conversation. XRDS: Crossroads, The ACM Magazine for Students, 2018.
[51] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J. of Global Optimization, 11(4):341–359, 1997.
[52] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 9446–9454, 2018.
[53] H. Wang, A. Deutz, T. Bäck, and M. Emmerich. Hypervolume indicator gradient ascent multi-objective optimization. In Evolutionary Multi-Criterion Optimization, volume 10173 of LNCS. Springer, 2017.
[54] Wikipedia. Casimir from l'ile aux enfants. https://en.wikipedia.org/wiki/L%27%C3%8Ele_aux_enfants#Characters, accessed 2021-01-18.
[55] L. Wolf, Y. Taigman, and A. Polyak. Unsupervised creation of parameterized avatars. In IEEE International Conference on Computer Vision (ICCV), pages 1530–1538, 2017.
[56] B. Xin, L. Chen, J. Chen, H. Ishibuchi, K. Hirota, and B. Liu. Interactive multiobjective optimization: A review of the state-of-the-art. IEEE Access, 6:41256–41279, 07 2018.
[57] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5485–5493, 2017.
[58] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In International Conference on Curves and Surfaces, pages 711–730. Springer, 2010.
[59] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 586–595, 2018.

[60] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. International Conference on Computer Vision (ICCV), 2017.
[61] S. Zhu, S. Fidler, R. Urtasun, D. Lin, and C. C. Loy. Be your own prada: Fashion synthesis with structural coherence. International Conference on Computer Vision (ICCV), 2017.

A Supplementary material

The following pages contains extended versions of some of the experiments described in the paper.



Frequency of non-trivial outputs per algorithm over the 12 settings.

Figure 6: Various experiments: the $(1+1)$-EA is best (as in [40]) when there is a risk of optimization-based artifacts (quality optimization), and MOO DE is best overall in other cases. **Top, similarity optimization on raw images:** MOO-optimizing the five similarity measures. We compare algorithms using the HV. The best methods, namely NGOpt12/9 (i.e. a MSR) and variants of DE based on MOO-specific operators, are actually not based on HV. **Middle, similarity optimization through PGAN** [25]: CVC. Average similarity (renormalized as detailed in the main text) for one of the similarity measures when MOO-optimizing the four other similarity measures (see text) with a latent representation. The MOO DE performs among the best for many values of the budget. **Bottom, quality optimization: CVC.** We have 12 contexts made of two settings and 6 budget values: we compute the best K512 (resp. Brisque in the second setting) score among the 16 images obtained by MOO of Blurriness and Brisque (resp. K512 in the second setting). We present the frequency at which non-trivial (finite) values were obtained over those 12 cases. This case is challenging, as the quality OFs are quite different from each other (Koncept512, based on a neural net for IQA; Brisque; and pure Blurriness) – for example, in many cases, we get a failure as Brisque value for the 16 output images of MOO optimizing Koncept512 and Blurriness, hence missing results. The only method that provided a non-trivial median result in all cases, consistently with results in [40], is (1+1)-EA: this algorithm, originally from the discrete optimization community, prefers optima with stable flat basins. Fischer's exact test: p-value $< 1.4\%$ for (1+1)-EA vs. any other.

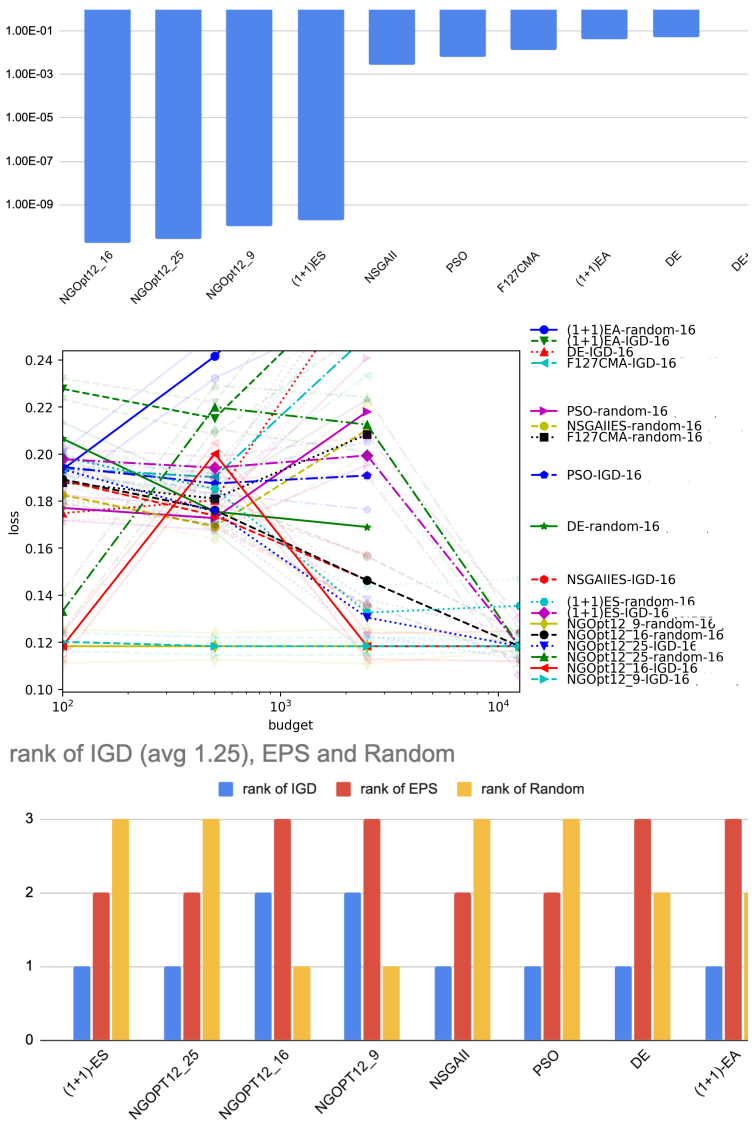


Figure 7: Settings close to inspirational generation (hence diversity is critical): MSR performing best, consistent with later experiments. **Many-objective similarity and quality optimization, through PGAN.** Images are represented by latent variables of a GAN. **Top, median HV** (log HV divided by max) obtained for each algorithm with budget 1e4: DE variants perform best, as in our artificial experiments. **Middle, comparison based on the CVC methodology:** best K512 score over images obtained by various MOO treatments of several similarity metrics and Brisque and Blurriness. We see that only MSR (and, interestingly, all variants of MSR) provides stable results in this inspirational generation context. This is consistent with previous results: whereas HV incorrectly predicts better results for DE. **Bottom: impact of the subsampling method** extracting M non-dominated points. For each algorithm, we compute the CVC loss, normalized and averaged. For each algorithm, we show the rank of the best result obtained by IGD (resp. EPS and Random) as subsampling methods. IGD significantly outperforms EPS (Wilcoxon p-value < 0.05).

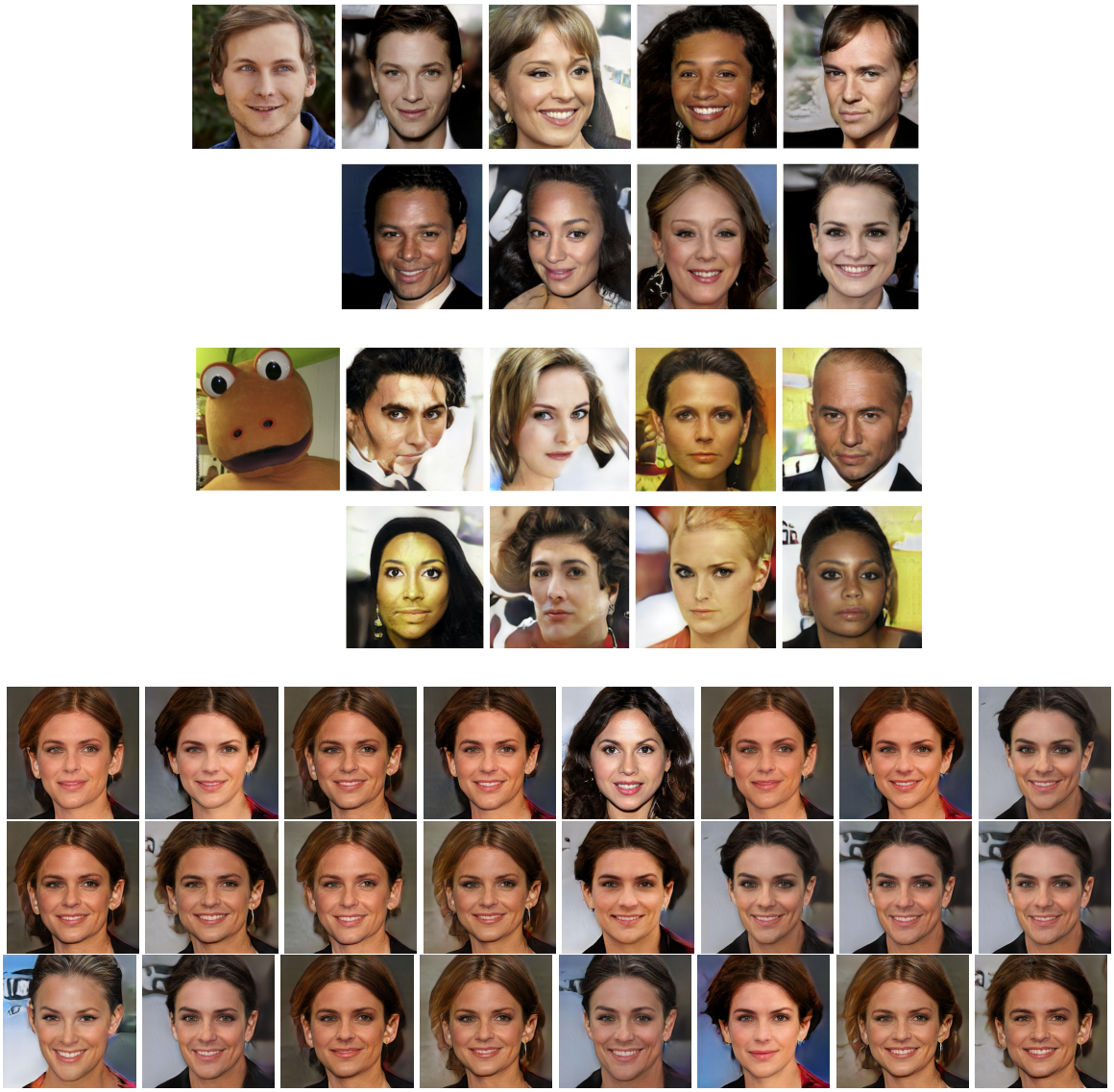


Figure 8: **MSRH for inspirational generation (projection onto the celebrities model): two examples in which (1) MOO optimization provides diversity (2) MOO by MSRH outperforms MOO by criteria using Pareto-dominance (3 versions).** In the context of inspirational generation (generating an image similar to a target), it turned out that the local optima of the original trade-off optimization from [40], obtained by repeated optimization runs (MSRH method), offer more diversity and quality than other multiobjective methods (see text: OF are valid only locally, leading to a diversity loss when applying Pareto-dominance globally, hence the success of MSR variants). MSR is already the best for CVC in Fig. 3. In these two cases above, the target image is the top left one. We present two hard cases, on which the original PytorchGanZoo code (without MOO) frequently fails, and PytorchGanZoo with Pareto-dominance generates very little diversity. **Top:** the original inspirational GAN tends to generate images of women whereas the target is male; in contrast to this, over the eight obtained images, our MOO code generates male faces. **Middle:** in this difficult case, we look for a face with similarities with Casimir [54], the gentle orange dinosaur. Two of the faces have the orange color and the big dark-circled white eyes. Three of the eight generated faces are complete failures, but with the diversity obtained by MOO the user can select the best of the eight generated images. **Bottom, unsuccessful inspirational generation using Pareto dominance, compared to top left:** we get very little diversity (all female). Conclusions: there is an intrinsic variance in inspirational generation so that MSRH does work quite well: using MSRH rather than applying Pareto-dominance means using objective functions only locally and avoids discarding dominated parts of the domain.

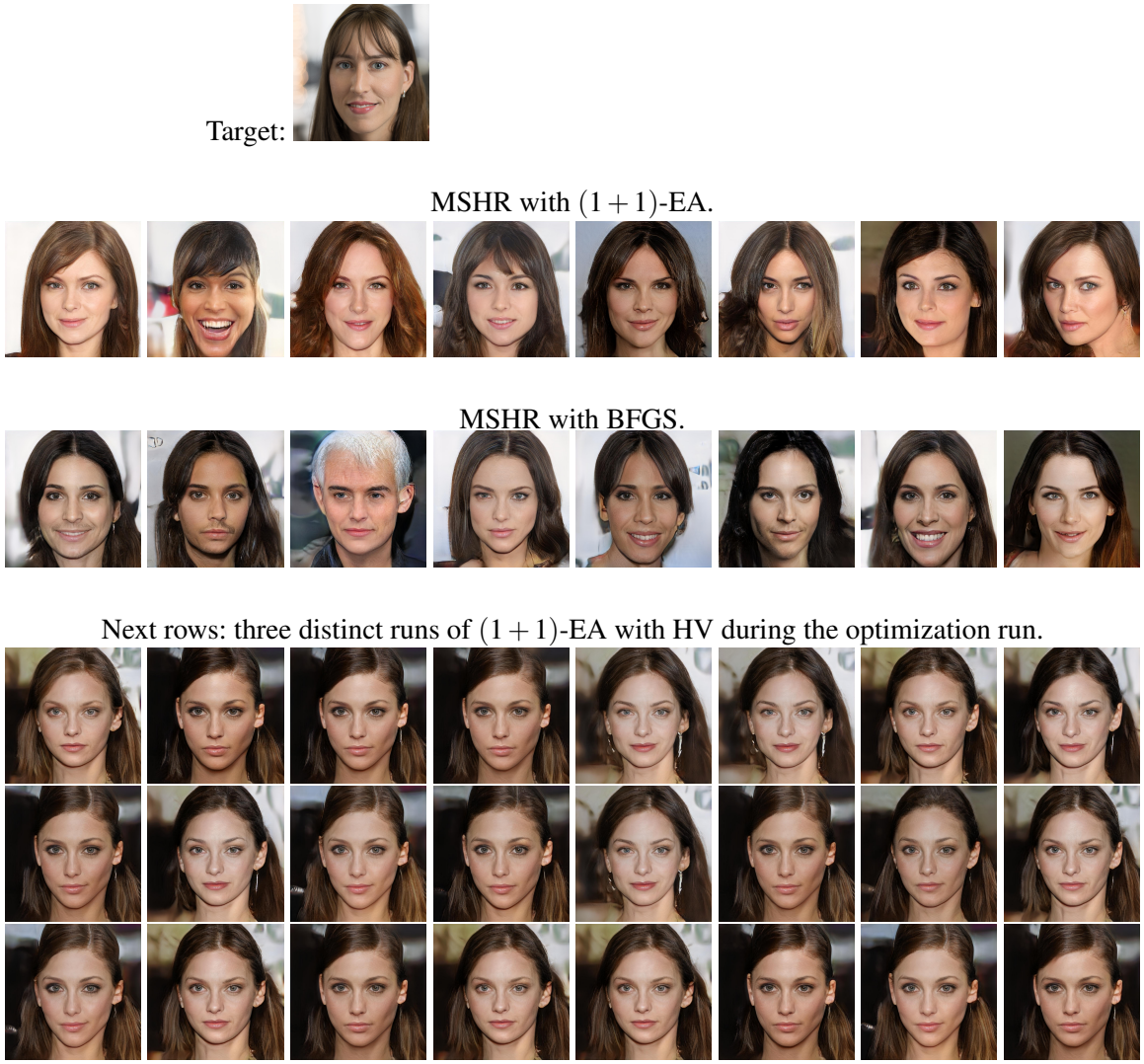
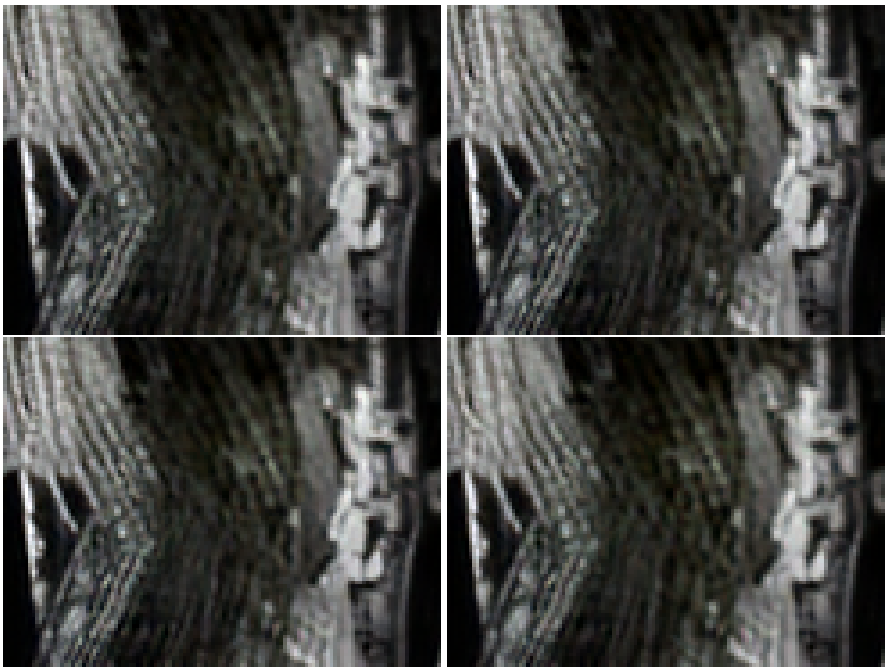


Figure 9: Other example, same setting as in Figure 4, top left and bottom: all images are projections of a same face. There is much more diversity in the two rows with MSHR: in contrast to Figure 4, all images of all methods satisfy the goal of projecting the target image (at least for part of their eight outputs) into the target domain.

DE



DCMA

Figure 10: **Comparison between MOO DE and DCMA (diagonal CMA optimizing HV).** Zooms on high-resolution images obtained by MOO-Tarsier using DE (a) and DCMA (b) on one image in Set14. In many cases, as in the present example, the images generated by DE are more diverse. For DE, the first image is blurrier and darker than the second one. The textures are all different for the four DE images.