Automated Feature Detection of Black-Box Continuous Search-Landscapes using Neural Image Recognition

Boris Yazmir
Migal Institute
Upper Galilee, Israel
borisy@migal.org.il

Ofer M. Shir
Tel-Hai College and Migal Institute
Upper Galilee, Israel
ofersh@telhai.ac.il

ABSTRACT

Extracting search-landscape features of high-dimensional Black-Box Optimization (BBO) problems, by minimal number of queries, is a grand challenge. Knowledge of such features per a given BBO problem-instance offers valuable information in light of the Algorithm Selection and/or Configuration problems. In practice, by solving this challenge, the query complexity of BBO may be reduced when capitalizing on algorithm portfolios. In this study we target the automated recognition of BBO benchmark functions. We propose an identification framework for d-dimensional continuous BBO test-suites, provided as input with a set of N search-points, sampled at random, together with their query values. We address it as a supervised multi-class image recognition problem, by introducing the concept of Landscape Images, and applying the basic LeNet5 Neural Network to classify them. The solution's core lies within the encapsulation of the BBO functions' data as Landscape Images, and the application of neural image recognition to learn their features. The efficacy of our approach is numerically validated on the noiseless COCO/BBOB test-functions, which are demonstrated to be correctly classified at high precision rates (≈90%) when N is in the order of d. This successful learning is another step toward automated feature detection of BBO problems.

CCS CONCEPTS

• **Computing methodologies** → **Randomized search**; *Continuous space search*; **Artificial intelligence**.

KEYWORDS

Black-Box Optimization, Algorithm Selection, Algorithm Configuration, noiseless COCO/BBOB, LeNet5 convolutional neural network

1 BACKGROUND AND CONTRIBUTION

The *algorithm selection problem* is defined as the task to choose the 'fittest' algorithm, within an available portfolio, per a given problem-instance. At the same time, the *algorithm configuration problem* refers to the task of choosing the 'fittest' parameter settings per a given problem-instance. Addressing those problems over *generalized problem features* [1] requires informative instance features, i.e., the problem of the learning model to choose the 'fittest' algorithm on uninformative instance features remains open. This open problem provides us with the motivation to obtain the ability to recognize BBO features and deduce local structure using a small number of queries. Our concrete contributions are the following:

(1) Encapsulating high-dimensional continuous BBO search-points within a so-called *landscape image* as an effective form of representing local information of the landscape.
(2) Translating the classification problem of such landscape images per their generating functions into an Image Recognition problem.
(3) Solving this problem using a convolutional neural network, and numerically validating it on the noiseless BBOB suite.

2 LANDSCAPE IMAGES' DEFINITION

Given an objective function, subject to minimization,

$$\varphi(\vec{x}^{(j)})_{k,i} : \mathbb{R}^d \to \mathbb{R}, \quad k = 1, 2, \ldots, F, \quad j = 1, 2, \ldots, N$$

let i denote a function instance (generated by, e.g., rotation and translation transformations on a base function), and let k denote the function type. N is the number of search-points $\vec{x}^{(j)}$. A *sample vector* is defined as the concatenation of a search-point and its associated objective function value:

$$\vec{\tau}^{(k,i,j)}(\vec{x}^{(j)}) := \left[\vec{x}^{(j)\,T}, \varphi_{k,i}\left(\vec{x}^{(j)}\right)\right] \in \mathbb{R}^{d+1}.$$

$M \times M$ is the landscape image size. Each image row is composed of $\vec{\tau}^{(k,i,j)}$ and $M - d - 1$ replications of $\varphi_{k,i}\left(\vec{x}^{(j)}\right)$. The first N rows correspond to the search-points. $M - N$ *zero vectors* and their function values $\varphi_{k,i}\left(\vec{0}\right)$ are then added to cope with possible underlying translations of $\varphi_{k,i}$:

$$\left\{\begin{array}{ll}
\vec{\tau}^{(k,i,1)}\left(\vec{x}^{(1)}\right)_{1\times d-1} & \varphi_{k,i}\left(\vec{x}^{(1)}\right)_{(M-d-1)\text{ replications}} \\
\cdots & \cdots \\
\cdots & \cdots \\
\vec{\tau}^{(k,i,N)}\left(\vec{x}^{(N)}\right)_{1\times d-1} & \varphi_{k,i}\left(\vec{x}^{(N)}\right)_{(M-d-1)\text{ replications}} \\
\vec{\tau}^{(k,i,N+1)}\left(\vec{0}\right)_{1\times d-1} & \varphi_{k,i}\left(\vec{0}\right)_{(M-d-1)\text{ replications}} \\
\cdots & \cdots \\
\cdots & \cdots \\
\vec{\tau}^{(k,i,M)}\left(\vec{0}\right)_{1\times d-1} & \varphi_{k,i}\left(\vec{0}\right)_{(M-d-1)\text{ replications}}
\end{array}\right\}$$

3 EXPERIMENTAL VALIDATION

We validated the aforementioned Image Recognition approach over the noiseless BBOB suite. We trained the basic LeNet5 model [2], over 32×32 images, for 24 classes per a spectrum of search-space dimensions $d \in \{2, 3, \ldots, 31\}$. Initially we accomplished learning

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3459439

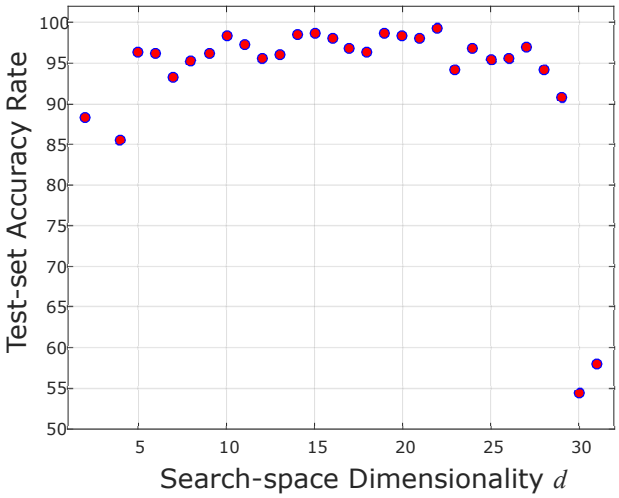

Figure 1: Test-set accuracy rates as a function of d

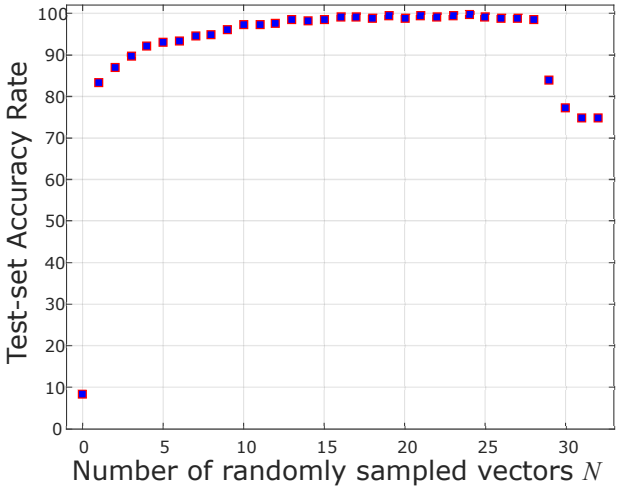

Figure 2: Test-set accuracy rates as a function of N per $d = 22$, to reflect the effect of increasing sampling versus padding with zero vectors.

over all dimensions d and explored the effectiveness of N. We trained this model over 3000 epochs with a learning rate of 10^{-6} and a batch size of 64, with one instance per function; the training and testing datasets were of sizes 24024 and 11952, respectively. We observed that the best accuracy rate was achieved for $d = 22$ at epoch 2987 (with accuracy rate 99.322%). Figure 1 depicts the best attained accuracy rates per each search-space dimension d. Furthermore, we explored the impact of N for the case of $d = 22$, and tested the model's performance when the number of sample vectors is increased, i.e. $N \in \{1, 3, \ldots, 32\}$. Following the aforementioned training setup, the best accuracy rate is observed for $N = 24$ and at epoch 2859 (accuracy rate 99.573%). Figure 2 depicts the best attained accuracy rates per each N when $d = 22$.

We also conducted experiments of the generalized multi-instance detection problem. Preliminary results of such experiments indicate

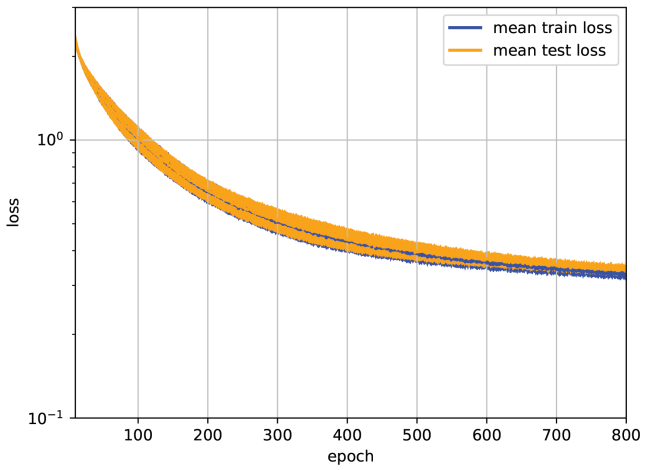

Figure 3: Averaged *loss* over 20 runs for the multi-instance detection problem at dimension $d = 22$ using Type-1 images with $N = 24$. LeNet5 is employed with a learning rate $\alpha = 10^{-5}$, batch size 64 and is run over 800 epochs. The training and testing datasets include 30120 and 14760 images, respectively. The standard deviation among runs is depicted as the shaded areas surrounding each curve.

that the generalized problem is learnable as well, but the existing configurations of LeNet5 obtain lower performance measures. For example, the learning process on 20 datasets with 5 possible instances per function is presented in Figure 3, depicting the *averaged loss* for this preliminary use-case. The averaged testing accuracy for the best nets trained on each of these 20 datasets was ~ 85%.

4 DISCUSSION

Evidently, the noiseless BBOB suit is learnable by a basic CNN. The neural learning model was capable of attaining efficient function recognition, using a carefully-designed database and a moderate number of learning iterations. The classification scores deteriorate only towards the end of the range. This behavior may be explained by the CNN's pooling effects, which diminish the influence of the least repeated data. For example, for $d = 30$, there are only 2 copies of sample vectors within our setup of $M = 32$, which can be lost due to pooling.

The preliminary multi-instance results show relatively high and robust recognition rates. These results can be improved using different landscape image structures and flexible CNN models.

This evident successful learning is another step toward automated feature detection and local structure deduction of BBO problems.

REFERENCES

[1] Tinus Abell, Yuri Malitsky, and Kevin Tierney. 2013. Features for Exploiting Black-Box Optimization Problem Structure. In *Revised Selected Papers of the 7th International Conference on Learning and Intelligent Optimization - Volume 7997* (Catania, Italy) *(LION 7)*. Springer-Verlag, Berlin, Heidelberg, 30–36. https://doi.org/10.1007/978-3-642-44973-4_4

[2] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*. 2278–2324.