DoE2Vec: deep-learning based features for exploratory landscape analysis

Stein, N. van; Long, F.; Frenzel, M.; Krause, P.; Gitterle, M.; Bäck, T.H.W.

DoE2Vec: Deep-learning Based Features for Exploratory Landscape Analysis

Bas van Stein*
LIACS, Leiden University
Leiden, The Netherlands
b.van.stein@liacs.leidenuniv.nl

Fu Xing Long
BMW Group
Munich, Germany
fu-xing.long@bmw.de

Moritz Frenzel
BMW Group
Munich, Germany
moritz.frenzel@bmw.de

Peter Krause
divis intelligent solutions GmbH
Dortmund, Germany
krause@divis-gmbh.de

Markus Gitterle
University of Applied Sciences
Munich, Germany
markus.gitterle@hm.edu

Thomas Bäck
LIACS, Leiden University
Leiden, The Netherlands
t.h.w.baeck@liacs.leidenuniv.nl

ABSTRACT

We propose DoE2Vec, a variational autoencoder (VAE)-based methodology to learn optimization landscape characteristics for downstream meta-learning tasks, e.g., automated selection of optimization algorithms. Principally, using large training data sets generated with a random function generator, DoE2Vec self-learns an informative latent representation for any design of experiments (DoE). Unlike the classical exploratory landscape analysis (ELA) method, our approach does not require any feature engineering and is easily applicable to high-dimensional search spaces. For validation, the proposed approach is used for three downstream classification tasks. We show that the latent representations can significantly boost performances when being used complementary to the classical ELA features.

CCS CONCEPTS

• **Computing methodologies → Continuous space search**; *Feature selection.*

KEYWORDS

exploratory landscape analysis, landscape features, auto-encoders, optimization

ACM Reference Format:
Bas van Stein, Fu Xing Long, Moritz Frenzel, Peter Krause, Markus Gitterle, and Thomas Bäck. 2023. DoE2Vec: Deep-learning Based Features for Exploratory Landscape Analysis . In *Genetic and Evolutionary Computation Conference Companion (GECCO '23 Companion), July 15–19, 2023, Lisbon, Portugal.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3583133.3590609

1 INTRODUCTION

Solving real-world black-box optimization problems can be extremely complicated, particularly if they are strongly nonlinear and require expensive function evaluations. As suggested by the no free lunch theorem, there is no such things as a single-best optimization algorithm, that is capable of optimally solving all kind of problems. The task in identifying the most time- and resource-efficient optimization algorithms for each specific problem, also known as the algorithm selection problem (ASP) (see [16]), is tedious and challenging, even with domain knowledge and experience. In recent years, landscape-aware algorithm selection has gained increasing attention from the research community, where the fitness landscape characteristics are exploited to explain the effectiveness of an algorithm across different problem instances (see [19, 21]). Beyond that, it has been shown that landscape characteristics are sufficiently informative in reliably predicting the performance of optimization algorithms, e.g., using machine learning approaches (see [1, 3, 5]). In other words, the expected performance of an optimization algorithm on an unseen problem can be estimated, once the corresponding landscape characteristics have been identified. Interested readers are referred to [6, 12].

Exploratory landscape analysis (ELA), for instance, considers six classes of expertly designed features, including y-distribution, level set, meta-model, local search, curvature and convexity, to numerically quantify the landscape complexity of an optimization problem, such as multimodality, global structure, separability, plateaus, etc. (see [10]). Each feature class consists of a set of features, which can be relatively cheaply computed. Other than typical ASP tasks, ELA has shown great potential in a wide variety of applications, such as understanding the underlying landscape of neural architecture search problems in [23] and classifying the black-box optimization benchmarking (BBOB) problems in [15]. Recently, ELA has been applied not only to analyze the landscape characteristics of crash-worthiness optimization problems from automotive industry, but also to identify appropriate cheap-to-evaluate functions as representative of the expensive real-world problems (see [9]). While ELA is well established in capturing the optimization landscape characteristics, we would like to raise our concerns regarding the following aspects.

(1) Many of the ELA features are highly correlated and redundant, particularly those within the same feature class (see [24]).
(2) Since ELA features are manually engineered by experts, their feature computation might be biased in capturing certain landscape characteristics (see [17]).

*Corresponding author

GECCO '23 Companion, July 15–19, 2023, Lisbon, Portugal

ACM ISBN 979-8-4007-0120-7/23/07.
https://doi.org/10.1145/3583133.3590609

(3) ELA features are less discriminative for high-dimensional problems (see [11]).

Instead of improving the ELA method directly, e.g., searching for more discriminative landscape features, we approach the problems from a different perspective. In this paper, we introduce an automated self-supervised representation learning approach to characterize optimization landscapes by exploiting information in the latent space. Essentially, a deep variational autoencoder (VAE) model is trained to extract an informative feature vector from a design of experiments (DoE), which is essentially a generic low-dimensional representation of the optimization landscape. Thus, the name of our approach: *DoE2Vec*. While the functionality of our approach is fully independent of ELA, experimental results reveal that its performance can be further improved when combined with ELA (and vice versa). To the best of our knowledge, a similar application approach with VAE in learning optimization landscape characteristics is still lacking. Section 2 briefly introduces the state-of-the-art representation learning of optimization landscapes as well as the concepts of (variational) autoencoder. This is followed by the description of our methodology in Section 3. Next, we explain and discuss our experimental results in Section 4. Lastly, conclusions and outlooks are included in Section 5.

2 REPRESENTATION OF OPTIMIZATION LANDSCAPE

In the conventional ELA approach, landscape features are computed primarily using a DoE of some samples points $\mathcal{W} = \{w_1, ..., w_n\}$ evaluated on an objective function f, i.e., $f: \mathbb{R}^d \rightarrow \mathbb{R}$, with $w_i \in \mathbb{R}^d$, n represents sample size, and d represents function dimension. The objective function values $f(w_i), i \in \{1, \ldots, n\}$ are the inputs of VAE models in DoE2Vec. In this work, we consider ELA features similar to those in [9], which do not require additional sampling, and compute them with the package `flacco` by [7]. These features include commonly used dimensionality reduction approaches such as principal component analysis (PCA), a number of simple surrogate models and many others.

To overcome the drawbacks of the ELA approach, attentions have been focused on developing algorithm selection approaches without requiring landscape features. For example, [13] proposed two feature-free approaches using a deep learning method, where optimization landscapes can be represented through either 1) image-based fitness maps or 2) graph-like fitness clouds. In the first approach, convolutional neural networks were employed to project data sets into two-dimensional fitness maps, using different dimensionality reduction techniques. In the second approach, data sets were embedded into point clouds using modified point cloud transformers, which can accurately capture the global landscape characteristics. Nonetheless, the fitness map approach suffered from the curse of dimensionality, while the fitness cloud approach was limited to fixed training sample size. Additional relevant works can be found in [14, 18]. Unlike these approaches, which were directly used as classifiers, the latent feature sets generated by our proposed approach can be easily combined with other features, such as ELA features, for classification tasks. In our work, we do not propose to replace conventional ELA features, but to actually extend them with autoencoder (AE) based latent-space features. Since the implementation of both approaches mentioned above is not available, a comparison to our work in terms of classifying high-level properties is only feasible by directly comparing their results on a identical experimental setup. Following this, results from the downstream tasks in this work can partially be compared to the mentioned results in [18], including the standard PCA, reduced multiple channel (rMC) and a transformer based approach (Transf.), taking into account that additional hyperparameter tuning was involved in their classification experiments with ELA features.

Our approach is capable of learning the representations of optimization landscapes in an automated, generic and unsupervised manner, with the advantage that the learned features are not biased towards any particular landscape characteristic. Unlike previously mentioned approaches, our proposed method is independent of the sampling method. By using only fully connected (dense) layers that learn from one-dimensional (flattened) landscapes, an AE or a VAE is, in theory, capable of learning any number of input-dimensions without scaling difficulties. Furthermore, the fast-to-train (V)AE models can be easily shared in practice.

2.1 (Variational) Autoencoder

A standard AE usually has a symmetrical architecture, consisting of three components: an encoder, a hidden layer, also known as bottleneck, and a decoder. In short, an encoder projects the input space $\mathcal{X}$ to a representative feature space $\mathcal{H}$, i.e., $e: \mathcal{X} \rightarrow \mathcal{H}$, while a decoder transforms the feature space back to the input space $d: \mathcal{H} \rightarrow \hat{\mathcal{X}}$ (see [2]). In other words, AE attempts to optimally reconstruct the original input space $\mathcal{X}$, by minimizing the reconstruction error $\mathcal{L}(\mathcal{X}, \hat{\mathcal{X}})$, e.g. mean squared error, during the (unsupervised) training process.

Unlike AE, the latent space of a Variational Auto Encoder (VAE) is encoded as a distribution by using a mean and variance layer, together with a sampling method. Following this, the latent space can be properly regularized to provide more meaningful features. Detailed explanations regarding VAE can be found in [8].

2.2 Black-Box Optimization Benchmarking

The development of DoE2Vec is based on the well-known academic BBOB suite by [4], consisting of altogther 24 noise-free real-parameter single objective optimization problems of different landscape complexity. For each BBOB problem, the global optimum (within $[-5, 5]^d$) can be randomly shifted, e.g., through translation or rotation, to generate a new problem instance.

3 DOE2VEC

Generally, our proposed method uses a VAE with an architecture of altogether seven fully connected layers, where rectified linear unit (ReLU) activation functions are assigned to the hidden layers, while a sigmoid activation function is used for the final output layer of the decoder. The encoder is composed of four fully connected layers with $dim(X)$ depending on the DoE sample size n, starting with the input layer size n, two hidden layers with sizes $n/2$ and $n/4$ and the latent size ls ($ls < n/4$) for the mean and log variance of the latent space. The decoder is composed of three fully connected layers with sizes $n/4$, $n/2$ and n for the final output layer. The focus of our approach lies on VAE, rather than AE, because it has the additional benefits of regularizing the latent space without loss of generalisation. For comparison, we consider a standard AE model as well (with the same number of hidden layers, except that the latent space is now a single dense layer, instead of a mean and log variance layer with a sampling scheme). Full specifications of the different models are available in our Github repository ([20]), while pre-trained models are also available on Huggingface.

The general workflow of DoE2Vec can be summarized as follows:

(1) First, a DoE of size 2^m is created, where m is a user defined parameter. By default, a Sobol sequence is used as sampling scheme, but any sampling scheme or even a custom DoE can be utilized in practice.
(2) The DoE samples, initially within the domain $[0, 1]^d$, can be re-scaled to any desired function boundaries, as the DoE2Vec method is scale-invariant by design.
(3) Next, the DoE samples are evaluated for a set of functions randomly generated using the random function generator from [9]. The main advantage of using this function generator is that a large training set can be easily created, covering a wide range of function landscape of different complexity.
(4) Following this, all objective function values are first re-scaled to $[0, 1]$ and then used as input vectors to train (V)AE models.
(5) Lastly, the latent space of the trained V(AE) models can be used as feature vectors for further downstream classification tasks, such as optimization algorithm selection.

In the next section, we will show that the learned feature representations have attractive characteristics and they are indeed useful for downstream classification and meta-learning tasks.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

In this work, we have conducted altogether four experiments for different research purposes, but only one experiment is included in this paper. For the remaining three experiments, that investigate the quality of the VAE models and the learned latent space, we refer to the full paper [22]. The experimental setup presented here includes three downstream multi-class classification tasks to show the potential of the proposed approach in practice, using the latent feature vectors as inputs.

In our experiments, we fix the sampling scheme to a Sobol sequence and m to eight, ending up with a DoE of 256 samples. Unless otherwise specified, all AE and VAE models are trained on a set of $250,000$ five-dimensional ($5d$) random functions.

The DoE2Vec approach is designed to learn characteristic representations of an optimization landscape, which can be verified through a classical classification task of high level function properties. These high level properties, such as multimodality, global structure and funnel structure, are important for the ASP, as they often determine the difficulty of an optimization problem. See [18] for the table that illustrates the BBOB functions and their associated high level properties.

In this experiment, a standard random forest (RF) model is implemented for the multiclass classification tasks based on the latent representations learned by four DoE2Vec models, consisting of AE-24, AE-32, VAE-24 and VAE-32. In other words, the high level properties of a BBOB function are to be predicted using the latent representations. Again, we compare the DoE2Vec approach against the classical ELA method, which is specifically constructed to excel in exactly this kind of function property classification tasks. Beyond that, a combination of classical ELA features with a VAE-32 model is included to evaluate the complimentary effect of the different feature sets.

The classification results (macro F1 scores) of the different feature sets are summarized in Table 1. It is not surprising that the ELA features generally perform very well and outperform the latent features most of the time, especially in classifying the global structure and multimodal landscapes. Fascinatingly, the classification performances can be significantly improved when the DoE2Vec is combined with the classical ELA method, indicating that both feature sets seem to be complimentary to each other.

5 CONCLUSIONS AND FUTURE WORK

In this work we propose *DoE2Vec*, a VAE-based approach to learn the latent representations of an optimization landscape, similar to the classical landscape features. Using a wide set of experiments (see [22]), we have shown that DoE2Vec is able to reconstruct a large set of functions accurately and that the approach can be used for downstream meta-learning tasks, such as algorithm selection. We provide an open-source documented implementation of our package at [20], with visualizations and video explanations. Pre-trained models (weights) are available on Huggingface. The proposed methodology can be effectively used next to existing techniques, such as classical ELA features, to further increase the classification accuracy of certain downstream tasks. In fact, DoE2Vec can learn good feature representations for optimization landscapes and has several advantages over the ELA approach, such as feature engineering or selection knowledge is not required, domain knowledge in ELA is not needed and it is applicable to optimization tasks in a very straightforward manner.

Nonetheless, there are a few known limitations to the proposed method, such as 1) our approach is scale-invariant, but not rotation- or translation-invariant. Using a different loss function to train the autoencoders might be able to improve this. 2) If a custom DoE sample is used, the model needs to be trained from scratch (no pre-trained model available). This typically takes a few minutes up to an hour, depending on the sample size n and number of random functions to train on. 3) The learned feature representations are a black-box that are hard to interpret directly. In future work, we plan to improve our approach by tackling some of the challenges mentioned.

ACKNOWLEDGMENTS

The contribution of this paper was written as part of the joint project newAIDE under the consortium leadership of BMW AG with the partners Altair Engineering GmbH, divis intelligent solutions GmbH, MSC Software GmbH, Technical University of Munich, TWT GmbH. The project is supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag.

REFERENCES

[1] Bernd Bischl, Olaf Mersmann, Heike Trautmann, and Mike Preuß. 2012. Algorithm Selection Based on Exploratory Landscape Analysis and Cost-Sensitive

Table 1: Classification results (averaged macro F1 scores over 10 runs with different random seeds) using a standard RF model with 100 trees, trained on the feature representations (from AE, VAE, classical ELA or ELA combined with VAE-32) of the first 100 instances for each BBOB function and validated on instance 101 to 120. * PCA, rMC and Transformer results are directly taken from the work of [18], which uses an identical experimental setup but without repetitions.

d	Task	AE-24	AE-32	VAE-24	VAE-32	ELA	PCA*	rMC*	Transformer*	ELA-VAE-32
2	multimodal	0.875	0.849	0.877	0.856	0.984	0.994	0.971	0.991	**0.991**
	global struct.	0.903	0.904	0.902	0.889	0.983	0.992	0.965	0.991	**0.998**
	funnel	0.985	0.974	0.956	0.978	**1.000**	0.999	0.995	1.000	**1.000**
5	multimodal	0.908	0.903	0.880	0.889	0.963	0.897	0.947	0.991	**0.998**
	global struct.	0.838	0.828	0.810	0.793	**1.000**	0.807	0.859	0.978	**1.000**
	funnel	**1.000**	**1.000**	0.996	0.991	**1.000**	0.990	0.989	1.000	**1.000**
10	multimodal	0.877	0.813	0.844	0.838	**1.000**	0.839	0.952	0.974	**1.000**
	global struct.	0.794	0.737	0.783	0.745	0.902	0.774	0.911	0.963	**0.991**
	funnel	0.998	0.993	0.997	0.993	0.972	0.977	0.991	**1.000**	0.997
20	multimodal	0.726	0.722	0.700	0.694	0.970	-	-	-	**0.991**
	global struct.	0.689	0.621	0.606	0.626	0.972	-	-	-	**0.997**
	funnel	0.993	0.982	0.985	0.982	**1.000**	-	-	-	**1.000**

Learning. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO '12)*. Association for Computing Machinery, New York, NY, USA, 313–320. https://doi.org/10.1145/2330163.2330209

[2] David Charte, Francisco Charte, María J. del Jesus, and Francisco Herrera. 2020. An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges. *Neurocomputing* 404 (sep 2020), 93–107. https://doi.org/10.1016/j.neucom.2020.04.057

[3] Johann Dréo, Carola Doerr, and Yann Semet. 2019. Coupling the Design of Benchmark with Algorithm in Landscape-Aware Solver Design. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '19)*. Association for Computing Machinery, New York, NY, USA, 1419–1420. https://doi.org/10.1145/3319619.3326821

[4] Nikolaus Hansen, Steffen Finck, Raymond Ros, and Anne Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions.* Research Report RR-6829. INRIA. https://hal.inria.fr/inria-00362633

[5] Anja Jankovic, Gorjan Popovski, Tome Eftimov, and Carola Doerr. 2021. *The Impact of Hyper-Parameter Tuning for Landscape-Aware Performance Regression and Algorithm Selection.* Association for Computing Machinery, New York, NY, USA, 687–696. https://doi.org/10.1145/3449639.3459406

[6] Pascal Kerschke and Heike Trautmann. 2019. Automated Algorithm Selection on Continuous Black-Box Problems by Combining Exploratory Landscape Analysis and Machine Learning. *Evolutionary Computation* 27, 1 (2019), 99–127. https://doi.org/10.1162/evco_a_00236

[7] Pascal Kerschke and Heike Trautmann. 2019. *Comprehensive Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems Using the R-Package Flacco.* Springer International Publishing, Cham, 93–123. https://doi.org/10.1007/978-3-030-25147-5_7

[8] Diederik P. Kingma and Max Welling. 2019. An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning* 12, 4 (2019), 307–392. https://doi.org/10.1561/2200000056

[9] Fu Xing Long, Bas van Stein, Moritz Frenzel, Peter Krause, Markus Gitterle, and Thomas Bäck. 2022. Learning the Characteristics of Engineering Optimization Problems with Applications in Automotive Crash. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Boston, Massachusetts) *(GECCO '22)*. Association for Computing Machinery, New York, NY, USA, 1227–1236. https://doi.org/10.1145/3512290.3528712

[10] Olaf Mersmann, Bernd Bischl, Heike Trautmann, Mike Preuss, Claus Weihs, and Günter Rudolph. 2011. Exploratory Landscape Analysis. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO '11)*. Association for Computing Machinery, New York, NY, USA, 829–836. https://doi.org/10.1145/2001576.2001690

[11] Mario A. Muñoz and Kate A. Smith-Miles. 2017. Performance Analysis of Continuous Black-Box Optimization Algorithms via Footprints in Instance Space. *Evol. Comput.* 25, 4 (dec 2017), 529–554. https://doi.org/10.1162/evco_a_00194

[12] Mario Andrés Muñoz, Yuan Sun, Michael Kirley, and Saman K. Halgamuge. 2015. Algorithm selection for black-box continuous optimization problems: A survey on methods and challenges. *Information Sciences* 317 (2015), 224–245. https://doi.org/10.1016/j.ins.2015.05.010

[13] Raphael Patrick Prager, Moritz Vinzent Seiler, Heike Trautmann, and Pascal Kerschke. 2022. Automated Algorithm Selection in Single-Objective Continuous Optimization: A Comparative Study of Deep Learning and Landscape Analysis Methods. In *Parallel Problem Solving from Nature – PPSN XVII*, Günter Rudolph, Anna V. Kononova, Hernán Aguirre, Pascal Kerschke, Gabriela Ochoa, and Tea Tušar (Eds.). Springer International Publishing, Cham, 3–17.

[14] Raphael Patrick Prager, Moritz Vinzent Seiler, Heike Trautmann, and Pascal Kerschke. 2021. Towards Feature-Free Automated Algorithm Selection for Single-Objective Continuous Black-Box Optimization. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. 1–8. https://doi.org/10.1109/SSCI50451.2021.9660174

[15] Quentin Renau, Johann Dreo, Carola Doerr, and Benjamin Doerr. 2021. Towards Explainable Exploratory Landscape Analysis: Extreme Feature Selection for Classifying BBOB Functions. In *Applications of Evolutionary Computation*, Pedro A. Castillo and Juan Luis Jiménez Laredo (Eds.). Springer International Publishing, Cham, 17–33. https://doi.org/10.1007/978-3-030-72699-7_2

[16] John R. Rice. 1976. The Algorithm Selection Problem. Advances in Computers, Vol. 15. Elsevier, 65–118. https://doi.org/10.1016/S0065-2458(08)60520-3

[17] Moritz Seiler, Janina Pohl, Jakob Bossek, Pascal Kerschke, and Heike Trautmann. 2020. Deep Learning as a Competitive Feature-Free Approach for Automated Algorithm Selection on the Traveling Salesperson Problem. https://doi.org/10.48550/ARXIV.2006.15968

[18] Moritz Vinzent Seiler, Raphael Patrick Prager, Pascal Kerschke, and Heike Trautmann. 2022. A Collection of Deep Learning-based Feature-Free Approaches for Characterizing Single-Objective Continuous Fitness Landscapes. arXiv:2204.05752 [cs.LG]

[19] David Simoncini, Sophie Barbe, Thomas Schiex, and Sébastien Verel. 2018. Fitness Landscape Analysis around the Optimum in Computational Protein Design. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '18)*. Association for Computing Machinery, New York, NY, USA, 355–362. https://doi.org/10.1145/3205455.3205626

[20] Bas van Stein. 2022. *doe2vec: paper release.* https://doi.org/10.5281/zenodo.7043301

[21] Bas van Stein, Michael Emmerich, and Zhiwei Yang. 2013. Fitness landscape analysis of nk landscapes and vehicle routing problems by expanded barrier trees. In *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation IV*. Springer, 75–89.

[22] Bas van Stein, Fu Xing Long, Moritz Frenzel, Peter Krause, Markus Gitterle, and Thomas Bäck. 2023. DoE2Vec: Deep-learning Based Features for Exploratory Landscape Analysis. arXiv:2304.01219 [math.OC]

[23] Bas van Stein, Hao Wang, and Thomas Bäck. 2020. Neural Network Design: Learning from Neural Architecture Search. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1341–1349. https://doi.org/10.1109/SSCI47803.2020.9308394

[24] Urban Škvorc, Tome Eftimov, and Peter Korošec. 2020. Understanding the problem space in single-objective numerical optimization using exploratory landscape analysis. *Applied Soft Computing* 90 (2020), 106138. https://doi.org/10.1016/j.asoc.2020.106138