Factors Impacting Diversity and Effectiveness of Evolved Modular Robots

FEDERICO PIGOZZI, University of Trieste, Italy
ERIC MEDVET, University of Trieste, Italy
ALBERTO BARTOLI, University of Trieste, Italy
MARCO ROCHELLI, University of Trieste, Italy

In many natural environments, different forms of living organisms successfully accomplish the same task while being diverse in shape and behavior. This biodiversity is what made life capable of adapting to disrupting changes. Being able to reproduce biodiversity in artificial agents, while still optimizing them for a particular task, might increase their applicability to scenarios where human response to unexpected changes is not possible. In this work, we focus on *Voxel-based Soft Robots* (VSRs), a form of robots that grants great freedom in the design of both morphology and controller and is hence promising in terms of biodiversity. We use *evolutionary computation* for optimizing, at the same time, morphology and controller of VSRs for the task of locomotion. We investigate experimentally whether three key factors—representation, Evolutionary Algorithm (EA), and environment—impact the emergence of biodiversity and if this occurs at the expense of effectiveness. We devise an automatic machine learning pipeline for systematically characterizing the morphology and behavior of robots resulting from the optimization process. We classify the robots into species and then measure biodiversity in populations of robots evolved in a multitude of conditions resulting from the combination of different morphology representations, controller representations, EAs, and environments. The experimental results suggest that, in general, EA and environment matter more than representation. We also propose a novel EA based on a speciation mechanism that operates on morphology and behavior descriptors and we show that it allows to jointly evolve morphology and controller of effective and diverse VSRs.

CCS Concepts: • **Computer systems organization** → **Evolutionary robotics**; • **Theory of computation** → *Evolutionary algorithms*; *Developmental representations*; • **Computing methodologies** → Mobile agents.

Additional Key Words and Phrases: Evolutionary robotics, Evolutionary algorithms, Neuroevolution, Representation, Diversity

1 INTRODUCTION

One grand-vision goal for robotics is to create robotic ecosystems that can endure hazardous and dynamic environments, pursuing their mission without any need for human supervision [Buchanan et al. 2020; Hale et al. 2019]. Several challenges must be overcome before achieving this goal, ranging from scalability in the design and building phases [Hale et al. 2020; Kriegman et al. 2020], possibly with auto-fabrication [Nitschke and Howard 2021], to effective mechanisms for robot adaptation to changes in both the environment and the robot itself (e.g., unforeseen damages) [Cully et al. 2015]. Adaptation would be a crucial property in order to ensure the survival of such autonomous robotic ecosystems.

The vast majority of works in robotics have aimed at finding one robot design that fits well for a given task in a given environment. One intrinsic limitation of this approach is that, when the task or the environmental

Authors' addresses: Federico Pigozzi, University of Trieste, Trieste, Italy, federico.pigozzi@phd.units.it; Eric Medvet, University of Trieste, Trieste, Italy, emedvet@units.it; Alberto Bartoli, University of Trieste, Trieste, Italy, bartoli.alberto@units.it; Marco Rochelli, University of Trieste, Trieste, Italy, marco.rochelli@gmail.com.

2688-3007/2023/3-ART $15.00
https://doi.org/10.1145/3587101

conditions change, the found design might become less effective, and possibly ineffective, making all the deployed robots immediately useless. That design, and the approach that produced it, are hence intrinsically not adaptable. Conversely, in nature, several different designs exist at the same time that fit well for a given task, e.g., locomotion, food harvesting, and reproduction. Diversity of designs for life forms, i.e., *biodiversity*, is hence the way nature achieves adaptation. Through biodiversity, natural evolution made life robust to disruptive changes, by filling a variety of ecological niches with different species. Indeed, biodiversity is so valuable that it has to be preserved in order to protect life itself [Tilman et al. 2017], as well as to increase the stability of ecosystems [Arese Lucini et al. 2020]. Diversity plays a fundamental role even in other settings as, e.g., in geology [Schrodt et al. 2019], economics [O'Sullivan and Sheffrin 2003], culture and politics [Young 1979].

Since researchers do resort to the paradigm of evolution for optimizing robots, by means of the methodologies of Evolutionary Robotics (ER) [Nolfi and Floreano 2000], it is sound to wonder if diversity may emerge from evolution also in populations of robots, similarly to what happens in populations of living creatures. In this work, we target this question and study which factors impact diversity in populations of evolved robots and whether they affect the effectiveness, i.e., the ability of robots to perform the required task.

Actually obtaining diversity while evolving robots is not an easy endeavor, though. First, comprehension, measurement, and promotion of diversity are important challenges themselves in the broader field of Evolutionary Computation (EC) [Črepinšek et al. 2013; Squillero and Tonda 2016], from which ER borrows the optimization techniques. Second, the characterization of diversity in ER and a deep understanding of which factors favor or impede it are still open issues [Silva et al. 2016]. In fact, the complexity of the robot-environment interplay makes it hard to obtain useful diversity, i.e., the diversity that does not affect effectiveness.

We consider the scenario of *Voxel-based Soft Robots* (VSRs), a kind of robots composed of cubic blocks of soft material capable of basic actuation and organized in a grid-like structure [Hiller and Lipson 2012]. VSRs fit particularly well our research questions for two reasons. First, their intrinsic modularity [Yim et al. 2007] and the fact they are an instance of soft robots [Kim et al. 2013] make VSRs a promising path toward autonomous robotics ecosystems. Second, they allow great freedom to the designers: in principle, they may optimize the morphology [Cheney et al. 2013], the controller [Talamini et al. 2019], and even the sensory apparatus [Ferigo et al. 2022a]. Indeed, VSRs have accomplished remarkable feats, e.g., squeezing through tight spaces [Cheney et al. 2015], crossing the sim-to-real gap [Kriegman et al. 2020], and igniting research on self-replicating synthetic organisms [Kriegman et al. 2021]. We investigate whether such great freedom can also foster diversity.

In this work, we study the impact of three factors on the diversity and effectiveness of populations of evolved VSRs: the solution representation, the Evolutionary Algorithm (EA), and the environment. To this end, we deal with the *joint* evolution of both morphology and controller of the robot, a task that is well-known to be problematic in the ER community [Lipson et al. 2016]. We consider two representations for the controller and two representations for the morphology that differ in their expressiveness and hence exhibit different potentials for diversity while allowing concurrent evolution of VSRs morphology and controller. We also consider four EAs that are radically different in how they deal with diversity and, finally, three environments with different degrees of difficulty. Among these, we propose and assess a novel EA that promotes diversity through speciation, where individuals are partitioned in species based on a few morphology and behavior descriptors that we designed for the case of VSRs doing locomotion.

A key contribution of our work is an automatic pipeline for *systematically* analyzing a very large number of VSRs (hundreds of thousands), while still looking at them with the human eye as we do when associating living organisms with a specific *species*. In doing so, we rely on Machine Learning (ML) to automatically assign species to VSRs. We build an ML pipeline for classifying VSRs into species according to morphology and behavior descriptors extracted from simulations of VSRs that perform the task of locomotion. Then, we use the relative abundance of predicted species as a measure of diversity for a population of VSRs, using the well-established Simpson index [Simpson 1949].

While there have been several studies addressing diversity in EC, with some of them considering the domain of robots, we believe that our work is the first to consider multiple factors affecting the diversity of both morphology and behavior. Moreover, we analyze those factors based on a notion of diversity that exploits humans' ability to discriminate between different approaches to locomotion, and hence facilitates comprehension. We believe our study may help future designers of evolvable robotic ecosystems in prioritizing different factors in terms of their impact on the diversity and effectiveness of evolved robots. Namely, we found that the environment and the EA seem to have a greater impact on diversity than the representation. Hence, these two factors should likely be the ones on which a designer should focus more.

This work is an extended version of [Medvet et al. 2021]. With respect to the cited work, the major additions are:

(a) we consider also the environment as a factor that can impact biodiversity and effectiveness, in addition to representation and EA;
(b) we consider two representations for the morphology, rather than the only one being used in [Medvet et al. 2021]: we also discuss the bias of the representation in terms of morphology descriptors;
(c) we increase the size of the robot morphology grid, from 5×5 to 10×10, in order to evolve larger and more complex robots and to better exploit the expressiveness of VSRs.

2 RELATED WORKS

There are earlier works that examined diversity in ER [Auerbach and Bongard 2014; De Carlo et al. 2020; Miras et al. 2020; Mouret and Doncieux 2012; Nordmoen et al. 2021; Samuelsen and Glette 2014]. Our work shares with them the general methodology, i.e., investigating the factors that may have an impact on evolution but differs in either the kind of robots considered or the factors themselves. Most of those works focused mainly on the impact of the environment [Auerbach and Bongard 2014; Gupta et al. 2021; Miras et al. 2020], while here we consider also the controller representation, the morphology representation, and the EA. The studies that are most similar to ours are [De Carlo et al. 2020] and [Mouret and Doncieux 2012]. The former investigates whether a mechanism of artificial speciation can favor morphological diversity. We also devise an EA that employs speciation and, in this respect, both the cited work and our approach were inspired by NEAT [Stanley and Miikkulainen 2002]. However, differently from [De Carlo et al. 2020], we (i) also consider the diversity of behavior, (ii) work with a more expressive kind of robots, VSRs, and (iii) analyze the joint impact of representation, EA, and environment. Mouret and Doncieux [2012] conducted an empirical study for comparing approaches for encouraging the diversity of the behavior of evolved robots. The authors focused mainly on the EA and on the measure of similarity on which to build diversity promotion. Different from the cited work, we (i) also consider the diversity of the robot morphologies and (ii) take the representation of solutions, i.e., how to map a genotype into a pair morphology-controller, as a factor potentially affecting diversity. We do not explicitly compare different similarity measures; yet, in the novel simple EA we propose for promoting diversity through speciation, we experiment with three different ways of measuring similarity between pairs of robots.

The pursuit of diversity has become more and more important in the EC community [Cully and Demiris 2017]. Proposals in this respect include novelty search [Lehman and Stanley 2008], novelty search with local competition [Lehman and Stanley 2011], and quality-diversity algorithms [Cully and Demiris 2017]. In a recent study, Nordmoen et al. [2021] showed that MAP-Elites, a form of quality-diversity optimization, is particularly suitable for exploiting the potential for diversity of modular (rigid) robots. The authors showed experimentally that populations of robots that evolved with MAP-Elites are more successful when transferred to new environments with respect to those that evolved with other EAs. Moreover, they specifically spotted a strong correlation between the diversity of earlier populations and effectiveness in locomotion in new environments. In a previous study, Tarapore et al. [2016] found that MAP-Elites is sensible to the representation of the controller in evolutionary

robotics applications: apparently, the lower locality of indirect, generative representations makes this EA less effective than with direct representations. Interestingly, MAP-Elites has also been recently found particularly effective in the simultaneous evolution of morphology and controller of VSRs [Ferigo et al. 2022b], with both indirect and direct representations of the controller. Overall, these results further motivate our aim of investigating the factors, beyond the EA, that favor or disfavor diversity in modular robotics.

Last but not least, our work fits into a relevant body of literature in ER concerning the joint evolution of morphology and controller. Past studies have either employed directed acyclic graphs [Sims 1994], L-systems [Hornby et al. 2001], gene regulatory networks [Joachimczak et al. 2016], direct encodings [Pagliuca and Nolfi 2020], or relied on more complex solutions as co-evolution [Cheney et al. 2018], and evolutionary reinforcement learning [Gupta et al. 2021]. In most of the cited works, the key ingredient for achieving the concurrent optimization of morphology and controller is the representation, i.e., how to encode in a genotype the information needed for describing both the morphology and the controller of the robot. While, in principle, some of the approaches mentioned above could be ported to the case of VSRs, in this study we "only" focus on four representations resulting from the combination of two for the morphology and two for the controller. Since we are interested in investigating the impact of representation (and its interplay with EA and environment) on diversity, we choose the representations for the morphology and the controller considering their compactness vs. expressiveness trade-off.

The intrinsic hardness of jointly evolving robots morphology and controller has been discussed by Lipson et al. [2016], using precisely the case study of VSRs. The reason for such difficulty rests on the *embodied cognition paradigm* [Pfeifer and Bongard 2006], which posits that intelligence emerges from the interaction between the controller (brain), the morphology (body), and the environment: brains evolve to fit a particular body and variations in the body are likely to cause mismatch [Eiben and Hart 2020]. On the other hand, allowing the concurrent optimization of both morphology and controller of the robots makes their optimization more difficult as diversity seems to vanish quickly [Pagliuca and Nolfi 2020]. An important contribution of our work is showing that it is possible to jointly evolve the morphology and the controller of VSRs that are effective in the task of locomotion, while also being diverse.

3 MATERIALS AND METHODS

3.1 Background

Voxel-based Soft Robots (VSRs) are a kind of robotic agents composed of several deformable cubes (called *voxels*), that perform actions by expanding or contracting the volume of the voxels. VSRs have been first introduced in [Hiller and Lipson 2012], along with a procedure for physically realizing them. We consider a 2-D variant of simulated (in discrete time and continuous space) VSRs, proposed in [Medvet et al. 2020b]. While disregarding one dimension makes these VSRs less realistic, it also reduces the search space; as such, this framework is particularly suitable for optimization by means of EC. We remark, however, that the representations and the algorithms proposed in this paper are easily portable to the 3-D setting.

In this section, we briefly describe the salient characteristics of VSRs that are relevant for this study: we refer the reader to [Medvet et al. 2020b,c] for more details.

A VSR is defined by its *morphology* and its *controller*. The former describes how the voxels are arranged and, for each voxel, the sensors it is equipped with. The controller determines how the area of each voxel varies over time, possibly relying on the readings of the sensors of that voxel and on communication with the other voxels. The ability of sensing both its internal state and the external environment makes VSRs potentially more effective in performing those tasks where sensing may be advantageous as, e.g., locomotion [Talamini et al. 2019].

3.1.1 Morphology. Voxels of a VSR are arranged in a 2-D grid topology of size $w \times h$. In our simulations, we model each voxel as the assembly of spring-damper systems, masses, and distance constraints [Medvet et al.

2020b]. Each voxel is rigidly welded to its four adjacent voxels (at the vertices), if present. We set the same values for the parameters of those components across all voxels, which results in all the voxels having the same mechanical properties.

The *area* of a voxel changes based on the corresponding actuation value, imposed by the controller, and on the external forces acting on the voxel, i.e., other voxels and the ground. The *actuation value* lies in $[-1, +1]$, where -1 corresponds to the maximum expansion and $+1$ corresponds to the maximum contraction. We model expansion and contraction in the simulation as instantaneous changes of the resting length of the springs in the spring-damper systems of the voxel.

Voxels can be equipped with sensors. Each sensor produces, at every time step, a *sensor reading* $\boldsymbol{s} \in \mathbb{R}^m$, where m is the dimensionality of the sensor type. We use three types of sensors and equip every voxel with one sensor of each type. *Area* sensors perceive the ratio between the current area of the voxel and its rest area (so $m = 1$). *Touch* sensors perceive whether the voxel is touching the ground or not ($m = 1$), and output 1 and 0, respectively. *Velocity* sensors perceive the velocity of the center of mass of the voxel along the x- and y-directions ($m = 2$). We apply a soft normalization (with tanh function) to every sensor reading, to ensure all sensor readings lie in $[0, 1]^m$. After normalization, to simulate real-world sensor noise, we perturb every sensor reading with additive Gaussian noise of mean 0 and standard deviation $\sigma_{\text{noise}} = 0.01$.

3.1.2 Controller. The controller determines the actuation value for each voxel at each time step. For the sake of this study, we resort to the distributed neural controller proposed in [Medvet et al. 2020a] that facilitates the joint evolution of morphology and controller. It consists of a number of fully connected, feed-forward Artificial Neural Networks (ANNs), one for every voxel, and operates as follows.

At time step $t = k\Delta t$, where Δt is the interval between two simulation time steps, each ANN (i) receives as input the local sensor readings and the $4n_{\text{comm}}$ *communication values* generated by the four adjacent ANNs at the previous time step and (ii) outputs the local actuation value and $4n_{\text{comm}}$ communication values to be fed to the adjacent ANNs at the next time step. Formally, each ANN works as follows:

$$\left[a^{(k)}\ \boldsymbol{o}_N^{(k)}\ \boldsymbol{o}_E^{(k)}\ \boldsymbol{o}_S^{(k)}\ \boldsymbol{o}_W^{(k)}\right] = \text{ANN}\left(\left[\boldsymbol{s}_i^{(k)}\ \boldsymbol{i}_N^{(k-1)}\ \boldsymbol{i}_E^{(k-1)}\ \boldsymbol{i}_S^{(k-1)}\ \boldsymbol{i}_W^{(k-1)}\right]\right), \tag{1}$$

where $a^{(k)} \in \mathbb{R}$ is the local actuation value, $\boldsymbol{o}_N^{(k)} \in \mathbb{R}^{n_{\text{comm}}}$ is the vector of communication values directed to the voxel above (similarly for E, S, W), $\boldsymbol{s}_i^{(k)} \in [0, 1]^4$ is the concatenation of the local sensor readings, and $\boldsymbol{i}_N^{(k-1)} \in \mathbb{R}^{n_{\text{comm}}}$ is the vector of communication values coming from the voxel above and been generated at the previous time step (similarly for E, S, W). If one of the neighbors is absent (e.g., if the voxel lies on the boundary of the morphology), we set $\boldsymbol{i}_N^{(k-1)}$ (or E, S, W) to a zero-vector $\mathbf{0}$ of the proper size. Similarly, we use a zero-vector as input communication values for all the voxels at the very first time step.

This form of controller meets the requirement of the concurrent evolution of morphology and controller: since the architecture of each ANN, namely the size of the input and output layers, is dictated only by the parameter n_{comm}, the structure of this distributed controller is agnostic with respect of the morphology of the VSR. Morphology and controller can hence be optimized together.

Despite its simplicity, this form of controller may result in interesting and variegate behaviors, since the interconnections between voxel ANNs make the overall architecture recurrent [Rumelhart et al. 1986], endowing the system with a form of "memory": therefore, there is a further dynamics introduced by the recurrent ANN that interacts with the dynamics induced by the mechanical model of the soft body (where the spring-and-hamper systems hold a form of "memory" too).

After some preliminary experiments and by taking into account the findings of [Medvet et al. 2020a], we use $n_{\text{comm}} = 1$, no hidden layers, and tanh as activation function for all the neurons—the latter guarantees that all communication values and the actuation value are in $[-1, 1]$. Considering that the overall dimension of the sensor

readings is 4, this results in each ANN having 4 + 4 input neurons and 4 + 1 output neurons; hence, each ANN is described by $8 \cdot 5 + 5 = 45 = n_{\text{ANN}}$ numerical parameters (the weights and biases of the edges connecting the neurons).

3.2 Measuring biodiversity

We take inspiration from the natural sciences and build our definition of biodiversity of populations of VSRs on the concept of *species*. A species groups together individuals sharing the same phenotypic traits. Once being able to associate each VSR with a species, we measure the diversity of a population of VSRs as the variety of species of its individuals.

For VSRs, since we deal with the joint evolution of morphology and controller, we use both kinds of traits for defining species. That is, we define a species as a pair of morphology species (based on morphology traits) and controller species (based on controller traits). Aiming at defining species that are based on *observable* traits, we actually consider the behavior, rather than the controller, of VSRs: while we do acknowledge that the behavior is not determined by the controller alone since it depends also on the morphology and the environment, we believe that the behavioral traits we chose to consider carry a significant amount of information about the underlying controller.

In order to have species that can be observed and discerned by humans, as it happens for living organisms, we *phenotypically classify* morphology species and behavior species by human categorization. Although it is just one of the ways that modern biology uses to define species, we believe this approach aligns with our objective of discovering species as seen by the human eye (rather than, e.g., according to genetic compatibility). However, since we evolve a large number of VSRs (in the order of hundreds of thousands), species classification by human inspection alone is not feasible. For this reason, we use ML for automatically determining the species of a VSR, an approach that is common also for determining the species of plants [Franklin and Ahmed 2018] and animals [Tabak et al. 2019].

We rely on *supervised* ML for species classification: we collect a few example cases, each one consisting of a VSR associated with a morphology class and a behavior class (both manually assigned); then, we learn two classification models, based on those examples, for associating any other VSR with one morphology and one behavior class. In order to use supervised ML, we first define the classes and a criterion for selecting a training set for manually assigning classes. We also define a set of descriptors (features, in ML terminology) useful for characterizing the morphology and the behavior of VSRs and to be fed to supervised learning techniques. In the following subsections, we describe each of the steps in detail.

3.3 Species classes definition and manual annotation

3.3.1 Classes definition. For deciding how many classes to use for morphology and behavior we observed a large number of videos of VSRs performing locomotion on three different terrains (flat, downhill, and uphill, see Section 4.3) and obtained through evolution with six different EAs (see Section 3.6.2), two different controller representations (see Section 3.6.1), and two different morphology representations (see Section 3.6.1).

Based on these observations, we define four classes for the morphology. *Blob* is compact and roundish, with no clear direction of development. *Limbed* has extrusions that might resemble limbs. *Elongated* is compact but with a clear direction of development. *Other* is a miscellaneous class for the VSRs that cannot be classified into one of the other classes. Figure 1 shows sample morphologies for the four classes.

Concerning the behavior, we define four classes. *Jumping* VSRs alternate between touching the ground and lifting their body up in the air. *Walking* VSRs cyclically alternate the parts of the body that touch the ground. *Rolling* VSRs roll on themselves. Finally, *Other* VSRs cannot be classified into one of the other labels, and thus fall

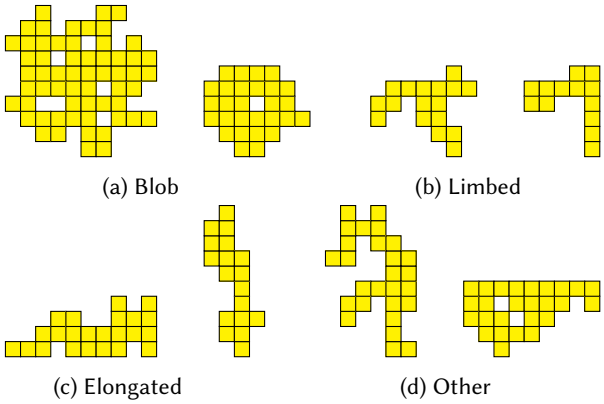


Fig. 1. Sample morphologies for the four morphology classes.

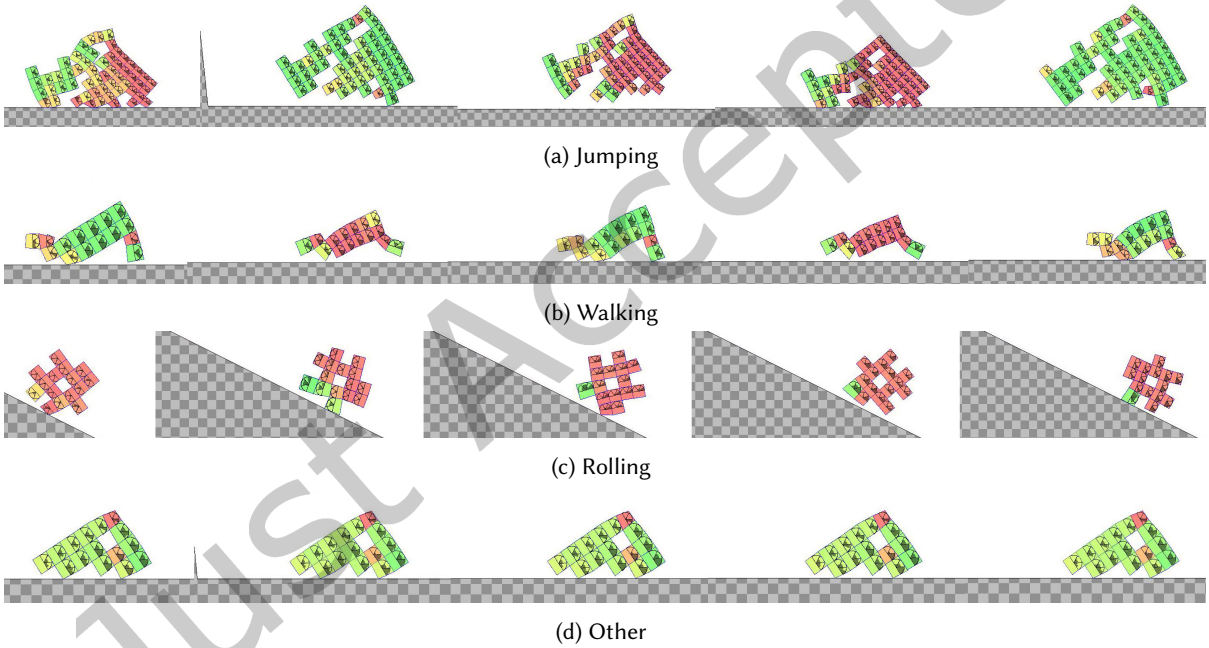


Fig. 2. Time-lapse showing locomotion for a sample VSR for each behavior class. The color of each voxel encodes the ratio between its current area and its rest area: red for < 1, yellow ≈ 1, green > 1; the circular sectors drawn at the center of each voxel indicate the current sensor readings.

in a miscellaneous class. Figure 2 show time-lapse images for the movement of a sample VSR for each behavior class. We also provide a video of those VSRs at https://youtu.be/tuD8scZ88Xc.

Since we define four classes for each one between morphology and behavior, and being the species the combination of the two, it follows that we define an overall number of 16 different species.

Table 1. Distribution of the manually assigned labels for the morphology and behavior classes on the training set.

	Blob	Limbed	Elongated	Other	Total
Jumping	0	1	291	25	317
Walking	67	82	64	174	388
Rolling	8	48	0	50	106
Other	82	93	132	437	744
Total	157	224	487	687	1555

3.3.2 Manual annotation. Having defined classes for morphology and behavior as illustrated in the previous section, we need to collect a training set of examples suitable for learning two classifiers that can automatically associate a previously unseen VSR with the corresponding classes.

Since we want to measure the diversity of populations of VSRs subjected to evolution, the training set should consist of individuals that uniformly cover the spaces of morphologies and behaviors that are likely to arise from evolution. However, a large portion of VSRs being generated during evolution (in particular at its early stage) perform poorly in the task of locomotion and are hence very difficult to associate with behavior class. For this reason, instead of simply selecting a random sample of all the individuals observed during the many evolutions we run (see Section 4.1, Section 4.2, and Section 4.3), we proceeded as follows.

First, we define three descriptors of the VSR morphology:

(1) *Number of voxels* d_{num}, i.e., the number of voxels in the VSR.
(2) *Elongation* d_{elong}, i.e., how stretched is the morphology in its direction of maximum development. For computing d_{elong} we first consider the smallest ellipse that encloses the VSR morphology; then we compute the ratio of the focal distance (distance between focal points) of the ellipse over the major axis length [Burger et al. 2009]. It follows that $d_{\text{elong}} \in [0, 1[$, with $d_{\text{elong}} = 0$ for perfectly "even" morphologies (e.g., a circle or square).
(3) *Compactness* d_{compact}, i.e., the ratio between the number of voxels of the VSR and those of the convex hull enclosing the morphology. The intuition is that d_{compact} is higher for morphologies with few concavities. Considering that the convex hull of a 2-D shape has an area that is always greater or equal than the area of the shape itself, it follows that $d_{\text{compact}} \in [0, 1]$, with $d_{\text{compact}} = 1$ for perfectly compact morphologies, i.e., morphologies with no concavities.

Then, we computed the values of d_{num}, d_{elong}, and d_{compact} for each VSR and partitioned the pool into 4 equal-size bins per descriptor, thus partitioning the pool of VSRs into $4 \times 4 \times 4 = 64$ bins. Next, we removed from all the 64 bins the VSRs that perform poorly in locomotion, i.e., those whose speed is lower than 2 m/s. Finally, we selected a subset of VSRs to be manually inspected for associating a morphology and a behavior class by taking randomly 25 VSRs out of each bin and adding 75 slow VSRs taken randomly from the discarded VSRs. This way, we assembled a set of $25 \cdot 64 + 75 = 1675$ VSRs.

We finally had a human operator inspect the VSRs, by looking at their unlabeled simulation videos, and assign VSRs to morphology and behavior classes until collecting at least 100 labels per class. We ended up with a training set of 1555 labeled VSRs, distributed among classes as summarized in Table 1.

The figures of Table 1 might suggest that our manual annotation procedure is affected by a human bias: human annotators tended to assign to the Other classes all the samples they were not able to assign to the clearly identifiable cases (both for the morphology and behavior). As a result, the dataset is slightly unbalanced and, as we will discuss in Section 4, this impacted later analyses. On the other hand, we believe that this bias is positive, in the sense that human operators did observe the robots in the context of a race for locomotion. In the end,

we are more interested in distinguishing between a crawling and a jumping artificial organism rather than a multitude of ways of being idle.

3.4 Features and learning

As in any ML application, two key design choices concern the features to extract for describing the observations (here, simulations of VSRs) and the learning technique. Since we aim at classifying the morphology and the behavior separately, we define different features for the two classification tasks. We describe them in the next subsections.

3.4.1 Morphology features. Since the VSR morphology is an arrangement of voxels in a 2-D grid, we could extract the features concerning the morphology directly from the grid, e.g., the descriptors described in Section 3.3. However, the grid is a static description of the VSR and does not capture the robot as seen during its life, i.e., as it could be seen by an external observer looking at the VSR while it does locomotion. For this reason, we define the morphology features based on the idea of the *dynamic pose* of the VSR. Intuitively, the dynamic pose can be regarded as a "long-exposure photograph" of the VSR during the simulation. We construct such a pose as follows.

Let a *snapshot* be the complete description of a time step of VSR simulation, i.e., it comprises the ground and every voxel of the VSR. Let S be a sequence of such snapshots, spanning an entire VSR simulation. For each snapshot, we determine the *minimal bounding square* $(x_0, y_0, x_0 + l, y_0 + l)$ around the VSR, that is, the smallest square parallel to the x-axis that completely encloses the VSR. Then, we partition the minimal bounding square in 16×16 inner squares with side length $\frac{l}{16}$ and build a matrix $\boldsymbol{d} \in \{0,1\}^{16\times16}$ where the element $d_{i,j}$ is 1 if and only if the corresponding inner square $(x_0 + (i-1)\frac{l}{16}, y_0 + (j-1)\frac{l}{16}, x_0 + i\frac{l}{16}, y_0 + j\frac{l}{16})$ is occupied by the VSR for at least half of the area. Finally, we compute the dynamic pose as the element-wise mode of the matrices computed for the snapshots in S.

We use the 256 values of the dynamic pose of a VSR as feature vector for its morphology, obtaining $\boldsymbol{f}_{\text{morph}} \in \{0,1\}^{256}$.

3.4.2 Behavior features. Since we deal with robots performing the task of locomotion, we define two groups of features that capture the VSR behavior while in locomotion, i.e., its *gait*, from two different points of view: the movement of the center of the VSR over the time and the way the VSR touches the ground while moving. We denote by $\boldsymbol{f}_{\text{center}}$ and $\boldsymbol{f}_{\text{footprints}}$ the two corresponding feature vectors, and by $\boldsymbol{f}_{\text{behavior}}$ their concatenation. We construct these vectors as follows.

Center movement. Concerning the features describing the movement of the center of the VSR, let S be a sequence of snapshots; we extract from S the discrete signals of the $x-$ and $y-$coordinate of the center of mass of the VSR. Then, we consider the signals of the first differences and compute their Fast Fourier Transform (FFT) [Cooley and Tukey 1965]. Subsequently, we take the magnitude of the two FFTs, filter out the components corresponding to frequencies greater than f_{max} (by taking into account the simulation time step Δt), and re-sample the remaining components in order to have n_{freq} components for each one of the two axes.

We use as feature vector $\boldsymbol{f}_{\text{center}}$ the concatenation of the two resulting vectors of magnitudes $\boldsymbol{f}_{\text{center}} = \left[\boldsymbol{f}_{\text{center},x}\ \boldsymbol{f}_{\text{center},y}\right] \in \mathbb{R}^{+2n_{\text{freq}}}$, with $\mathbb{R}^+ = [0, +\infty[$. After preliminary experiments and leveraging our expertise, we set $f_{\text{max}} = 10\,\text{Hz}$ and $n_{\text{freq}} = 100$.

Footprints. Concerning the features describing how the VSR touches the ground, we build a definition based on the concept of *footprint*. Given a snapshot, we consider the projection $[x_0, x_0 + l]$ of the minimal bounding square on the x-axis and we partition it in 8 equally-sized segments. Then we build the footprint of the VSR in that snapshot as a binary sequence $\boldsymbol{m} \in \{0,1\}^8$, where the element m_i is 1 if and only if the VSR is touching the ground for at least half of the corresponding segment $[x_0 + (i-1)\frac{l}{8}, x_0 + i\frac{l}{8}]$.

Given a sequence S of snapshots, we follow the following procedure to determine the set of footprint features. (1) We split S in a sequence $(S_1, S_2, \dots)$ of non-overlapping subsequences, each one corresponding to an interval of $\Delta t_{\text{footprint}}$ simulated time (we set $\Delta t_{\text{footprint}} = 0.5$ s after some preliminary experiments). (2) We build the sequence $M = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \dots\}$ of footprints where each $\boldsymbol{m}_i$ is obtained as the element-wise mode of the footprints computed from snapshots in S_i. (3) We consider all the non-overlapping n-grams of footprints in M, with $2 \leq n \leq 10$, that occur at least twice and compute the overall duration of each n-gram, computed as the product between its number of occurrences and its duration. (4) We select as the *main footprint n-gram* $M^\star$ the n-gram with the greatest overall duration. (5) We compute the following features for $M^\star$: duration $|M^\star|\Delta t_{\text{footprint}}$, average touch area $\frac{1}{|M^\star|}\frac{1}{8}\sum_{\boldsymbol{m}\in M^\star}\sum_{i=1}^{i=8} m_i$, number of occurrences of $M^\star$ in M, mode $\Delta t_{M^\star}$ of the intervals between subsequent occurrences of $M^\star$, rate of intervals that are equals to the mode.

We use as feature vector the five features computed for the main footprint n-gram $M^\star$, i.e., $\boldsymbol{f}_{\text{footprint}} \in \mathbb{R}^{+5}$.

3.4.3 Learning technique. We rely on Random Forest [Breiman 2001] as classifiers for the morphology and behavior classes based on $\boldsymbol{f}_{\text{morph}}$ and $\boldsymbol{f}_{\text{behavior}}$, respectively. We chose this supervised learning technique because studies [Fernández-Delgado et al. 2014; Wainberg et al. 2016] have proved it to be among the best general purpose classification techniques. We used the default values for the main parameters: 100 trees in the ensemble and $m = \lfloor\sqrt{p}\rfloor$ features (i.e., $m = \lfloor\sqrt{|f_{\text{morph}}|}\rfloor = 16$ for the morphology classifier and $m = \left\lfloor\sqrt{|f_{\text{behavior}}|}\right\rfloor = 14$) for each tree.

For an estimate of the accuracy of Random Forest on our two classification tasks, we performed a 5-fold cross validation assessment using the 1555 labeled simulations (Table 1) and obtained an average accuracy of 0.833 and 0.891 for morphology and behavior classification, respectively. A trivial classifier (always predicting the most frequent class) taking into account the class imbalances obtained an accuracy of 0.442 and 0.478, respectively.

3.5 Simpson index

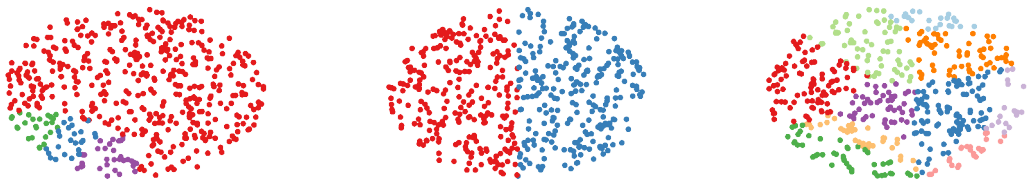
Different measures of diversity have been used in the ecological literature [Magurran 2013]. Among them, the Simpson index is one of the most commonly used [Simpson 1949]. Given a population of individuals that is partitioned based on species, this index is defined as $\lambda = \sum_{i=1}^{i=n} p_i^2$, where n is the number of species and p_i is the fraction of individuals of the i-th species.

Intuitively, the Simpson index measures the probability that two individuals picked at random with replacement belong to the same species. Since its semantics is the opposite of the one of diversity (i.e., $\lambda = 1$ for a population composed of a single species and it is < 1 for more diverse populations), in this study we use the Inverse Simpson index (ISI) λ^{-1} (defined in $[1, +\infty[$): the greater the ISI, the more diverse the population. Since we defined a limited number of possible species, i.e., 16, the actual domain of ISI in our study is $[1, 16]$.

Simpson index is a suitable measure of diversity since it depends on both the total number of species, as well as their relative abundance. To better grasp this intuition, Figure 3 plots three example populations colored by species and their corresponding ISI. As can be seen, the red species pollutes most of the population in Figure 3a, and, as a result, the corresponding ISI is the lowest. The population in Figure 3b is evenly partitioned between two species; nevertheless, there are just two of them, so the total variety of species is not very high, and indeed ISI is only slightly higher than before. Finally, the population in Figure 3c witnesses a great variety of species, and all of them are evenly balanced in terms of abundance, which is reflected in its ISI.

3.6 Evolution of VSR morphology *and* controller

We want to investigate how different factors (namely, the representation, the EA, and the environment) impact effectiveness and diversity in populations of VSRs, i.e., whether VSRs can be optimized for a given task, by means



(a) ISI = 1.41 (b) ISI = 2.00 (c) ISI = 7.82

Fig. 3. Example populations colored by species, and their corresponding ISI.

of EC, while maintaining diversity measured as above. For fully exploiting the potential of expressing diverse solutions to a task, we need a way for evolving simultaneously the morphology and the controller.

We here propose different genotypic representations that jointly encode a description of both the morphology and the controller of a VSR in a single numerical vector $\boldsymbol{v} \in \mathbb{R}^p$. The resulting optimization problem is hence a search in the numerical space $\mathbb{R}^p$, for which many techniques do exist. We experiment with two representations for the controller and two for the morphology. We also experiment with four EAs that fit this scenario, two of them being tailored to the specific goal of promoting diversity. In the following subsections, we describe the representations and the EAs.

3.6.1 Representations. We define two representations for the controller and two representations for the morphology of the VSR in the form of a numerical vector $\boldsymbol{v} \in \mathbb{R}^p$. In both of them, a portion $\boldsymbol{v}_{\text{morph}}$ of $\boldsymbol{v}$ encodes a description of the morphology and the remaining, disjoint portion $\boldsymbol{v}_{\text{ctrl}}$ describes the controller, i.e., $\boldsymbol{v} = \left[\boldsymbol{v}_{\text{morph}}\ \boldsymbol{v}_{\text{ctrl}}\right]$. The controller representations differ in the latter, whereas the morphology representations differ in the former.

Controller. We propose two alternatives for the controller defined in Section 3.1.2. In the *Homogeneous* controller representation, denoted by Ho, we assume that all the ANNs have the same parameters $\boldsymbol{w}$. It follows that $\boldsymbol{v}_{\text{ctrl}} = \boldsymbol{w} \in \mathbb{R}^{n_{\text{ANN}}}$.

In the *Heterogeneous* controller representation, that we denote by He, we assume that ANNs may have different parameters and that $n_{\text{size}} \times n_{\text{size}}$ is the size of the largest representable VSR. In order to favor the locality of the representation [Rothlauf 2006] and to make the controller representation agnostic with respect to the morphology representation, $\boldsymbol{v}_{\text{ctrl}}$ is the concatenation of the weights of the ANNs of all the voxels, i.e., $\boldsymbol{v}_{\text{ctrl}} = \left[\boldsymbol{w}_{1,1}\ \ldots\ \boldsymbol{w}_{n_{\text{size}} \times n_{\text{size}}}\right]$, where $\boldsymbol{w}_{i,j}$ is the vector of parameters of the ANN at the (i, j) position in the enclosing grid. It follows that $\boldsymbol{v}_{\text{ctrl}} \in \mathbb{R}^{n_{\text{size}}^2 n_{\text{ANN}}}$. This representation is the same, for the controller part, as the one proposed in [Medvet et al. 2020a].

The two controller representations differ in expressiveness. The heterogeneous representation is the most expressive one, thus resulting in the largest search space. The homogeneous representation is the least expressive one: its search space is smaller and hence, in principle, easier to explore. However, it might be harder for evolution to find the combination of genes that, when translated to the same ANNs for each voxel, results in a VSR that exhibits the desired complex behavior.

Morphology. We propose two alternatives for representing the morphology. In the *Direct* morphology representation, we associate each gene with one and only one voxel of the final morphology, i.e., a collection of adjacent voxels arranged in a 2-D grid. Given a $\boldsymbol{v}_{\text{morph}} \in \mathbb{R}^{n_{\text{size}}^2}$, we build a morphology as follows. Let $n_{\text{size}} \times n_{\text{size}}$ be the size of a square grid enclosing the largest representable VSR morphology. First, we build a Boolean matrix $\boldsymbol{b} = \{\text{T}, \text{F}\}^{n_{\text{size}} \times n_{\text{size}}}$ where $b_{x,y}$ is set to true if and only if $v_k > 0$, with $k = x + (y-1)n_{\text{size}}$. Then, we build the

morphology by considering the largest connected component of $\boldsymbol{b}$ elements set to true and putting a voxel at each element of such set. As a consequence, $\boldsymbol{v}_{\text{morph}}$ comprises one number for every voxel in the grid, i.e., $|\boldsymbol{v}_{\text{morph}}| = n_{\text{size}} \times n_{\text{size}}$. Albeit simple, such direct representations have proved effective for the joint evolution of morphology and control of other kinds of embodied agents [Ha 2019; Pagliuca and Nolfi 2020]. Moreover, the irregularities that may arise from this direct representation of a VSR morphology have been shown to be potentially beneficial for the adaptability of morphologies to different tasks [Talamini et al. 2021].

In the *Gaussian Mixture Model (GMM)* morphology representation, based on the observations of [Cheney et al. 2013; Hiller and Lipson 2012], we use the generative representation based on a mixture of bi-variate Gaussian distributions [Lindsay 1995] described in [Medvet et al. 2020c]. Let n_{GMM} be the number of Gaussians in the mixture. First, we build a real matrix $\boldsymbol{b} = \{\text{T}, \text{F}\}^{n_{\text{size}} \times n_{\text{size}}}$ where $b_{x,y}$ is set to true if and only if $f(x', y') > 0$, with:

$$f(x', y') = \sum_{i=1}^{n_{\text{GMM}}} \frac{\phi_i}{2\pi\sigma_{i,x}\sigma_{i,y}} \exp^{-\frac{1}{2}\left(\frac{(x'-\mu_{i,x})^2}{\sigma_{i,x}} + \frac{(y'-\mu_{i,y})^2}{\sigma_{i,y}}\right)},$$

where $x' = 2\frac{x}{n_{\text{size}}} - 1$ and $y' = 2\frac{y}{n_{\text{size}}} - 1$ are the x, y coordinates normalized in $[-1, 1]$, $\boldsymbol{\mu_i} = [\mu_{i,x}, \mu_{i,y}]$ and $\boldsymbol{\sigma}_i = \begin{bmatrix} \sigma_{i,x} & 0 \\ 0 & \sigma_{i,y} \end{bmatrix}$ are the mean vector and the covariance matrix for the i-th Gaussian, and $\phi_i \in [0, 1]$ is its mixing coefficient. Then, we build the morphology by considering the largest connected component of $\boldsymbol{b}$ elements set to true and putting a voxel at each element of such set. Since we restrict every $\boldsymbol{\sigma}_i$ to be diagonal, $\boldsymbol{v}_{\text{morph}}$ comprises 5 numbers for every Gaussian in the mixture, i.e., the two means, the two variances, and the mixing coefficient, so that $\boldsymbol{v}_{\text{morph}} \in \mathbb{R}^{5n_{\text{GMM}}}$. Note that we clip the values of $\boldsymbol{v}_{\text{morph}}$ corresponding to $\sigma_{i,x}$ and $\sigma_{i,y}$, for all i, in order to make them positive—e.g., for the first matrix, we set $\sigma_{1,x} = \max(0, v_{\text{morph},3})$.

The two morphology representations differ along two axes. First, they differ in terms of the compactness-expressiveness trade-off: the Direct representation presents, in general, a larger search space, thus holding the potential for more expressiveness. The GMM representation is potentially more compact (depending on the actual value of n_{GMM}), thus holding the potential for easier exploration of the search space. Second, we confront a direct representation with an indirect (or generative) one. In doing so, we tap ourselves into the debate surrounding direct and indirect encodings in ER [Veenstra et al. 2017]. As a consequence of those differences, the two representations potentially differ in the types of morphologies they are more suited to encoding. On one side, we expect the Direct representation to encode more irregular morphologies; while [Cheney et al. 2013] proved direct representations to be sub-optimal for the evolution of VSR morphologies, they can increase the degree of complexity of a dynamical system [Talamini et al. 2021] and put it in a better position to exploit cognitive offloading, i.e., moving from the brain to the body the ability to store and processing information [Nolfi 2021]. On the other side, we expect the GMM representation to generate morphologies that are more regular, symmetrical, and composed of a few limbs. We evaluate how the morphology representations differ along these three axes by looking into the bias of the representation in Section 4.1.2.

For the sake of this study, we performed the experiments with $n_{\text{size}} = 10$ and $n_{\text{GMM}} = 5$, resulting in $|\boldsymbol{v}_{\text{ctrl}}|$ being $45 \cdot 10 \cdot 10 = 4500$ and 45, respectively for He and Ho representations, and $|\boldsymbol{v}_{\text{morph}}|$ being $10 \cdot 10 = 100$ and $5 \cdot 5 = 25$, respectively for Direct and GMM representations.

3.6.2 Evolutionary algorithms. We use four EAs suitable for optimizing in the numerical space $\mathbb{R}^p$. Two of them are general purpose EAs, one is an EA that employs a form of speciation aimed at favoring diversity in the population—yet not based explicitly on the concept of species defined in Section 3.3—and one is a quality-diversity algorithm, a family of approaches that aim at returning a population that is both diverse as possible and effective as possible.

Evolution strategy. The first EA is Canonical-ES (ES) [Chrabaszcz et al. 2018], a state-of-the-art Evolution Strategy. Evolution Strategies [Beyer and Schwefel 2002; Schwefel 1965] constitute a family of EAs including some variants that have recently shown to achieve competitive results for continuous control tasks and game-playing [Salimans et al. 2017]. ESs have also been used for evolving the controller of VSRs [Ferigo et al. 2021; Nadizar et al. 2021].

ES iteratively evolves a fixed-size population of n_{pop} individuals as realizations of a multivariate normal distribution of mean $\boldsymbol{\mu} \in \mathbb{R}^p$ that is updated during the evolution. At iteration, n_{pop} children are born from $\boldsymbol{\mu}$, each one obtained by applying Gaussian noise $\boldsymbol{\epsilon}_i$ with $\sigma_{\text{mut}} = 0.35$ and zero mean:

$$\boldsymbol{v}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i \tag{2}$$

where $\boldsymbol{v}_i$ is the i-th child. Then, we update $\boldsymbol{\mu}$ by selecting the fittest quarter of the children and correcting $\boldsymbol{\mu}$ according to a weighted mean of the corresponding $\boldsymbol{\epsilon}_i$:

$$\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \sum_{i=1}^{i=\left\lfloor \frac{n_{\text{pop}}}{4} \right\rfloor} w_i \boldsymbol{\epsilon}_i \tag{3}$$

with weights w_i set as in [Hansen and Ostermeier 1996]:

$$w_i = \frac{\log\left(n_{\text{pop}} + 0.5\right) - \log i}{\sum_{j=1}^{j=n_{\text{pop}}} \log\left(n_{\text{pop}} + 0.5\right) - \log j} \tag{4}$$

We set $\boldsymbol{\mu}$ by sampling uniformly in the interval $[-1, 1]$ for each vector element. We set $n_{\text{pop}} = 40$ and let ES iterate until $n_{\text{evals}} = 30\,000$ fitness evaluations have happened.

ES is a form of population-based optimization. We remark, however, that the population in ES is indeed a realization of a multivariate normal distribution, i.e., all the individuals are "variations" of a single individual, the mean of the distribution. This observation is relevant in our settings, where we study the diversity of the evolved solutions.

Genetic algorithm. As a second EA, we use a standard variant of Genetic Algorithm (GA). Our GA variant iteratively evolves a fixed-size population of n_{pop} individuals according to a $\mu + \lambda$ generational model [De Jong 2006], i.e., with overlapping: at each generation, the offspring and the parents are merged and the worst half individuals are discarded. For building the offspring, we select individuals with tournament selection of size 5 and then apply Gaussian mutation with $\sigma_{\text{mut}} = 0.35$, with probability p_{mut}, or extended geometric crossover with probability $1 - p_{\text{mut}}$. For extended geometric crossover, given two parents $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^p$, the new individual is born as $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{\alpha}(\boldsymbol{v}_2 - \boldsymbol{v}_1)$, where each element α_i of $\boldsymbol{\alpha}$ is chosen randomly with uniform probability in $[-1, 2]$. In this way, the new individual may fall outside the hypercube defined by the parents, hence favoring exploration. Moreover, we perturb each child of crossover by applying Gaussian mutation with $\sigma_{\text{mut}} = 0.1$, to prevent having genetically identical children based on selecting similar parents.

As for ES, we build the initial population by sampling uniformly in the interval $[-1, 1]$ and iterate until $n_{\text{evals}} = 30\,000$ fitness evaluations have happened. Moreover, we set $n_{\text{pop}} = 100$ and $p_{\text{mut}} = 0.2$.

Speciated evolver. We designed this EA, which denote by SE (for Speciated Evolver), specifically for this study. SE employs a form of speciation inspired by NEAT [Stanley and Miikkulainen 2002], the popular EA for evolving the topology and the weights of ANNs. NEAT employed speciation with the purpose of protecting innovations introduced by modifications in the topology. In SE, we do not optimize the topology of the ANNs composing the controller of the VSR, while we do optimize the morphology of the VSR. Our goal is hence not to protect innovation, but explicitly to promote diversity.

1 **function** evolve():
2 $P \leftarrow$ initialize(n_{pop})
3 **while** ¬shouldStop() **do**
4 $P' \leftarrow \emptyset$
5 $(P_1, \dots, P_k) \leftarrow$ kmeans(P)
6 $P' \leftarrow P' \cup \{\text{best}(P)\}$
7 **foreach** $i \in \{1, \dots, n\}$ **do**
8 **if** $|P_i| \geq n_{\text{elite}}$ **then**
9 $P' \leftarrow P' \cup \{\text{best}(P_i)\}$
10 **end**
11 **end**
12 $n'_{\text{pop}} \leftarrow n_{\text{pop}} - |P'|$
13 $\boldsymbol{r} \leftarrow \text{ranks}(\text{repr}(P_1), \dots, \text{repr}(P_k))$
14 **foreach** $i \in \{1, \dots, k\}$ **do**
15 $c \leftarrow 0$
16 **while** $c < n'_{\text{pop}} \alpha^{r_i} \frac{1}{\sum_{i=1}^{i=k} \alpha^{r_i}}$ **do**
17 **if** $U(0,1) \leq p_{\text{mut}}$ **then**
18 $\boldsymbol{v} = \text{nth}(P_i, c \bmod |P_i|)$
19 $P' \leftarrow P' \cup \{\text{mutate}(\boldsymbol{v})\}$
20 $c \leftarrow c + 1$
21 **else**
22 $\boldsymbol{v}_1 = \text{nth}(P_i, c \bmod |P_i|)$
23 $\boldsymbol{v}_2 = \text{nth}(P_i, (c+1) \bmod |P_i|)$
24 $P' \leftarrow P' \cup \{\text{crossover}(\boldsymbol{v}_1, \boldsymbol{v}_2)\}$
25 $c \leftarrow c + 2$
26 **end**
27 **end**
28 **end**
29 $P \leftarrow P'$
30 **end**
31 **end**

Algorithm 1: The algorithm of SE.

Similarly to ES and GA, SE iteratively evolves a fixed-size population of n_{pop} individuals, as shown in Algorithm 1. At each iteration, individuals are partitioned in species according to a given criterion (described below) that also elects a single representative individual of each species (lines 5 and 13). Then, the current best individual in the population and the best individual of every species larger than n_{elite} are moved in the offspring (lines 6–11). The remaining individuals in the offspring are generated as follows. First, an offspring slot of size $n'_{\text{pop}} \alpha^{r_i} \frac{1}{\sum_{i=1}^{i=n} \alpha^{r_i}}$ is reserved to each species P_i depending on the rank r_i of the corresponding representative individual $\text{repr}(P_i)$ (line 16)—$\alpha \in]0, 1]$ is a parameter of the algorithm, the closer to 1, the less the preference for fittest species. Then, the offspring slot is filled by applying Gaussian mutation or expanded geometric crossover (as in GA) to individuals of the corresponding species P_i (lines 17–25).

We explore three variants of SE. All three of them use the k-means clustering technique [Lloyd 1982] for partitioning the population into species and elect as representative individual the one closest to the centroid of

the cluster. They differ in the features that are used for clustering individuals into species. In the first variant, which we denote by SE-g, we use the genotype $\boldsymbol{v}$, whose dimension depends on the representation. In the second variant, denoted by SE-s, we use the vector $\boldsymbol{f}_{\text{morph}} \in \mathbb{R}^{256}$ of the morphology features (see Section 3.4.1). Finally, in the third variant, denoted by SE-b, we use the vector $\boldsymbol{f}_{\text{behavior}} \in \mathbb{R}^{205}$ of the behavior features (see Section 3.4.2). In all cases, we compute the Euclidean distance after having properly normalized the vectors of the individuals of the current population.

We designed SE with the main goal of studying the factors affecting diversity; the search effectiveness (both in terms of fitness and diversity) was not a design goal, while we believe that SE simplicity favors the analysis and the interpretation of the experimental results. We remark, however, that approaches similar to SE have been proposed and successfully employed in the past for evolving robots, in particular using behavioral similarity for speciating solutions [Trujillo et al. 2011].

In the experiments, we set $n_{\text{pop}} = 100$, $p_{\text{mut}} = 0.2$, as for GA, and iterate until $n_{\text{evals}} = 30\,000$ have happened, as for GA and ES. Moreover, we set $\alpha = 0.75$, $n_{\text{elite}} = 5$, and $k = 10$ (for k-means).

Quality-diversity algorithm. Finally, we experiment with an established quality-diversity algorithm, Multidimensional Archive of Phenotypic Elites (MAP-Elites, hence further abbreviated as ME) [Cully et al. 2015].

ME computes a descriptor $\boldsymbol{d} \in \mathbb{R}^q$ for every individual to partition the descriptor space in a grid of n_{bin} cells for every dimension. ME starts with a population of n_{parents} individuals randomly initialized in $[-1, 1]^p$, evaluates them, computes their descriptors, maps the descriptors to the corresponding cell in the grid, and selects the best performing individual of every non-empty cell. These individuals form the archive. At each iteration, n_{parents} children are born by mutating n_{parents} randomly chosen individuals in the archive with Gaussian mutation with $\sigma_{\text{mut}} = 0.35$, they are evaluated, and their descriptors are computed; if the descriptor of a child maps to a cell that is empty or stores an individual of lower fitness, the child is added to the archive and the individual of lower fitness is discarded. The algorithm iterates until n_{evals} fitness evaluations have happened.

We used the morphology descriptors of Section 3.3.2, so $\boldsymbol{d} = [d_{\text{num}}, d_{\text{elong}}, d_{\text{compact}}] \in \mathbb{R}^3$. We set $n_{\text{bins}} = 10$, $n_{\text{parents}} = 20$, and $n_{\text{evals}} = 30\,000$. This choices resulted in a grid with 10^3 cells.

ME is a simple yet effective algorithm that intrinsically creates an incentive to fill as many cells in the grid as possible, thus covering as much of the descriptor space as possible. Moreover, it has also been recently found particularly effective in the simultaneous evolution of morphology and controller of VSRs [Ferigo et al. 2022b]. Here, we employed a simple and widespread version of ME, as we did for the other EAs considered in this study: however, later improvements have been proposed for ME which further increased its effectiveness, as, e.g., the directional variation operator proposed by Vassiliades and Mouret [2018].

4 EXPERIMENTS AND DISCUSSION

We aim at investigating how the three key factors of representation (both controller and morphology), EA, and environment impact diversity and effectiveness.

For all the experiments in the following sections, we considered the task of locomotion. The goal of the VSR is to travel as fast as possible, in the positive x direction, on a flat surface and within a time interval of $t_{\text{final}} = 30$ s (simulated time). The fitness of the individual is the average velocity v_x, measured considering the position of the center of mass of the VSR at the beginning and end of the simulation:

$$v_x = \frac{x_c(t_{\text{final}}) - x_c(0)}{t_{\text{final}}} \tag{5}$$

where $x_c(t)$ is the x-position of the center of mass of the VSR at time t. We remark that each simulation of any given VSR is deterministic.

Table 2. Summary of the experimental parameters.

Context	Name	Description	Value
Controller	σ_{noise}	Standard deviation of the additive Gaussian noise applied to sensor readings	0.01
Controller	n_{comm}	Number of communication values between adjacent voxels	1
Representation (all)	n_{size}	Side length of the largest representable VSR	10
Representation (GMM)	n_{GMM}	Number of Gaussian distributions in the mixture	5
EA (ES)	σ_{mut}	Standard deviation of the Gaussian noise applied to children	0.35
EA (ES)	n_{pop}	Population size	30
EA (GA, SE-*)	p_{mut}	Probability of applying only mutation	0.2
EA (GA, SE-*, ME)	σ_{mut}	Standard deviation of the Gaussian mutation	0.35
EA (GA, SE-*)	n_{pop}	Population size	100
EA (SE-*)	α	Species preference	0.75
EA (SE-*)	n_{elite}	Minimum species size to preserve elite individual	5
EA (SE-*)	k	Number of species (found by k-means)	10
EA (ME)	n_{bin}	Number of bins per descriptor dimension	10
EA (ME)	n_{parents}	Number of randomly chosen parents	20
EA, all	n_{evals}	Number of fitness evaluations (as stop criterion)	30 000
Simulation	t_{final}	Duration of a simulation for the locomotion task (in s)	10
Simulation	Δt	Simulation time step (in s)	$\frac{1}{60}$

We used 2D-VSR-Sim [Medvet et al. 2020b] for the simulation of the VSRs, with a time step of $\Delta t = \frac{1}{60}$ s and all the other parameters set to default values. The code of the experiments is publicly available at https://github.com/pigozzif/VSRBiodiversity. Table 2 reports an overview of all the parameter values used in our experiments.

For each experiment, i.e., a combination of representation, EA, and environment, we performed 10 evolutionary runs by varying the random seed of the EA. Table 3 summarizes, for each factor of investigation, the experimental settings. During each experiment, we saved the entire population of VSRs every 1000 fitness evaluations (to meet our storage constraints), which resulted in 3000 individuals for every run.

As a result, $10 \cdot 4 \cdot 2 \cdot 3000 + 10 \cdot 2 \cdot 3120 + 10 \cdot 2 \cdot 2 \cdot 3000 + 10 \cdot 2 \cdot 2 \cdot 3000 + 10 \cdot 3000 = 572\,400$ VSRs were generated in all of our experiments: for each one of them we applied the two classifiers for predicting the morphology and behavior classes learned on the subset of 1555 manually labeled VSRs, as described in Section 3.2.

We carried out all statistical tests with the two-sided Mann-Whitney U rank test for independent samples, using, unless otherwise specified, 0.05 as the confidence level.

We consider average velocity $v_x^\star$ of the best individual as a measure of effectiveness and ISI (see Section 3.2) as a measure of diversity. We inquire into the landscape of diversity in the population of robots by taking the median values (across evolutionary runs) of $v_x^\star$ and ISI at the last generation. To gain more insights, for some experiments we also present the number of VSRs broken-down by classes, at the last generation, and the distribution of morphology (or behavior) descriptors.

Every experimental configuration was able to evolve effective VSRs for the task of locomotion. We manually inspected a subset of the most effective VSRs (across every combination of representation/EA/environment tested) and found that they looked quite different. We showcase some of those VSRs in Figure 4; the corresponding video can be found at https://youtu.be/_kblILsfivw. Those hand-picked VSRs strikingly mirror emergent patterns found

Table 3. Summary of the experiments. The table shows one row for each factor potentially impacting diversity and effectiveness. For each factor, the table shows the different experiments we performed, i.e., the different combinations of controller representation, morphology, EA, and environment.

		Representation				
Factor	Section	Contr.	Morph.	EA	Env.	N. of exp.
Contr. repr.	4.1.1	Ho, He	Direct	GA, SE-s	Flat	4
Morph. repr.	4.1.2	Ho	Direct, GMM	GA, SE-s	Flat	4
EA	4.2	Ho	Direct	ES, GA, SE-g, SE-s, SE-b, ME	Flat	6
Environment	4.3	Ho	Direct	GA, SE-s	Flat, Downhill, Uphill	6

in nature. We nicknamed "gecko" one individual (Figure 4g), for example, for its clinging to an inclined surface and climbing up with a pair of short limbs. Another individual, "bigfoot" (Figure 4d), walked by treading a big extrusion that looked like a foot. Others slithered like centipedes or trotted like equines, to name a few traits. In general, both primitive and complex morphologies emerged. Interestingly, evolution succeeded in adaptation also with bizarre and unusual solutions. In fact, some individuals covered long distances while possessing a morphology that might have turned out a handicap. As a proof of concept, one individual (nicknamed the "snail", Figure 4i) crawled forward despite carrying on its back an uncomfortable hump (resembling a shell indeed) that might have hindered motion.

4.1 Impact of the representation

In the next subsection, we investigate the impact of the representation on effectiveness and diversity for both the controller and the morphology.

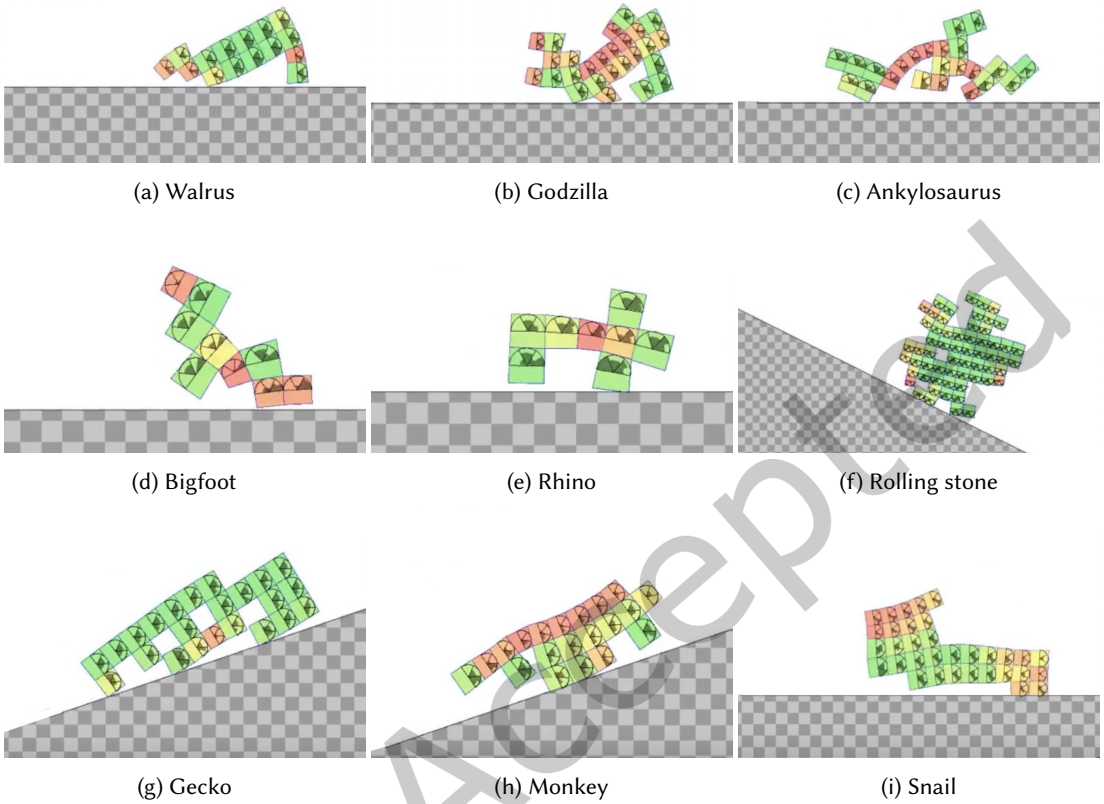
4.1.1 Controller representation. We performed an experimental campaign of 10 evolutionary runs on the He and Ho controller representations using GA and SE-s (i.e., the variant of SE in which the partitioning criterion is based on morphology features). We chose these two EAs because, as it turned out from our experiments described in Section 4.2, they were the ones with the weakest ability to promote diversity (GA) and strongest ability to promote effectiveness (SE-s). Moreover, for the sake of this experiment, we adopt Direct as the sole morphology representation and discuss the comparison with GMM in the next subsection.

We report the results in Figure 5 in terms of $v_x^\star$ and ISI at the last generation, together with the p-values for every EA.

From the plot, two interesting conclusions can be made:

(a) Ho outperforms He with both EAs in terms of effectiveness, and
(b) there is no clear difference in terms of diversity.

The former means that despite the lower expressiveness of the Ho representation, a "single ANN" (i.e., one whose weights and biases are shared among all the voxels) is capable of driving cooperatively an entire robot when proper parameters are found. As it turns out from our experiments, finding these parameters is feasible with both EAs, probably because of the much lower dimension of the search space. p-values are significant for both EAs. Regarding conclusion (b), the much smaller search space induced by Ho does not result in a lower diversity when compared to He. The two representations are indeed comparable in terms of ISI; neither of the p-values is significant.



(a) Walrus (b) Godzilla (c) Ankylosaurus

(d) Bigfoot (e) Rhino (f) Rolling stone

(g) Gecko (h) Monkey (i) Snail

Fig. 4. A subset of outperforming individuals, in terms of effectiveness and variety of morphologies and behaviors. A video can be found at https://youtu.be/_kblILsfivw.

To conclude, the compactness of the search space—championed by Ho—triumphs over the expressiveness of the representation—championed by He. For these reasons, we adopt Ho as the only representation for the experiments in the next sections.

As an aside, we note that there is a contrast in terms of diversity between the two EAs, regardless of the controller representation (Figure 5, right): we discuss this aspect in more detail in Section 4.2.

Figure 6 reports, for each species resulting from the combination of a morphology class and a behavior class, the rate (bubble size) of VSRs at the last generation belonging to that species and the average velocity v_x (bubble color) of the best VSR of that species (that is not, in general, the best of the entire population).

While there seems to be no clear difference with SE-s, we spot some trends with GA. Diversity for Ho is greater along the behavior axis, with a relative majority of Walking individuals. On the other side, He seems to favor more diversity along the morphology axis, with Limbed and Other individuals being equally represented. Overall, in all the cases the percentage of robots following in the Other class for both morphology and behavior is significant: we remark that this finding may be explained partly in terms of the dataset used for training the classifier, which is slightly unbalanced (see Section 3.3.2).

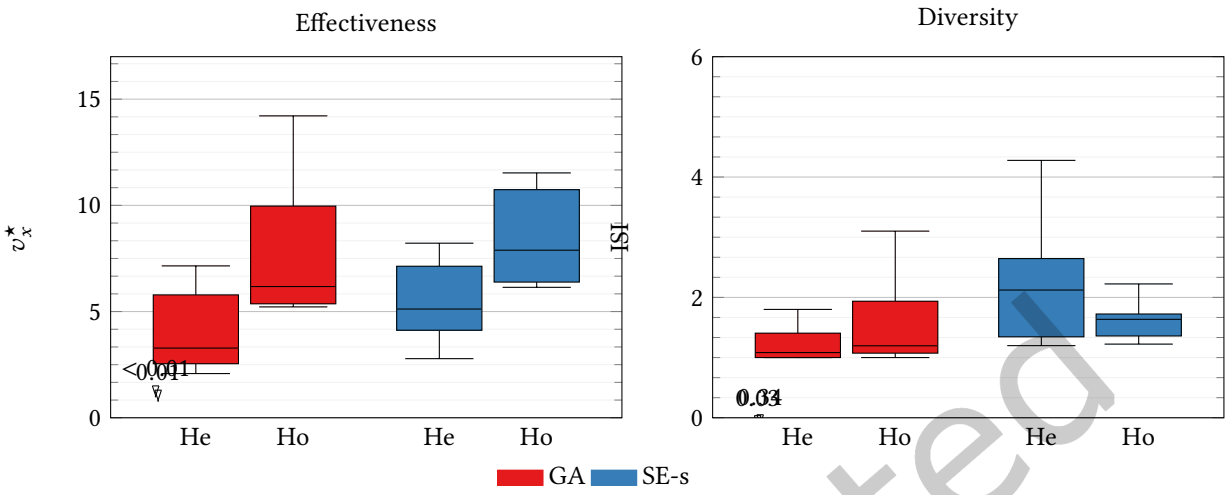


Fig. 5. Boxplots of $v_x^\star$ and ISI of the population at the last generation, obtained with four combinations of EA and controller representation (10 evolutionary runs for each combination). Each lower (upper) whisker is at the smallest (larger) data value greater than the lower (upper) quartile − (+) 1.5IQR, IQR being the inter quartile range. Numbers above each pair of boxes are p-values. The Ho representation outperforms the He one in terms of fitness, while there is no clear difference in terms of diversity.

4.1.2 Morphology representation. We aim at investigating what impact the two morphology representations, namely Direct and GMM, have on effectiveness and diversity. As discussed in Section 4.1.1, we adopt Ho as controller representation, GA and SE-s as EAs. With these settings, we performed an experimental campaign of 10 evolutionary runs with the Direct and GMM morphology representations.

We report the results in Figure 7 in terms of $v_x^\star$ and ISI at the last generation, together with the p-values for every EA.

From Figure 7, it clearly turns out that the Direct and GMM representations are comparable in terms of effectiveness, with no clear differences across the EAs and no significant p-values. Considering diversity, GMM performs better with SE-s, and the p-value is significant. Once again, we remark that there seems to be a significant effect of the EA on diversity, and we treat it in Section 4.2.

We observe that both morphology representations are capable of evolving effective individuals, as well as preserving a fair amount of diversity (if we employ the appropriate EA). Thus, considering also the results of Section 4.1.1, we speculate that joint optimization of morphology and control is more susceptible to the choice of the controller rather than morphology representation. One reason might be that the morphology representations we consider in this work do not significantly differ in the dimension of the search space, whereas the same is not true for the controller representations.

We thus asked ourselves whether the two representations have a different bias in terms of species and emergent forms of "life", and so what classes (i.e., species) did prevail. Figure 8 presents the breakdown by classes with the same visual syntax of Figure 6.

As can be seen from Figure 8, there are relevant differences in the GA case. In particular, the Direct representation is more capable of evolving individuals in the Walking behavior class.

Both representations witness a proliferation of individuals in the Other behavior class, which consists, for the most part, of idle individuals. This result confirms the previous observation that the joint evolution of morphology and controller is a difficult optimization problem: the majority of offspring individuals are born with brains that

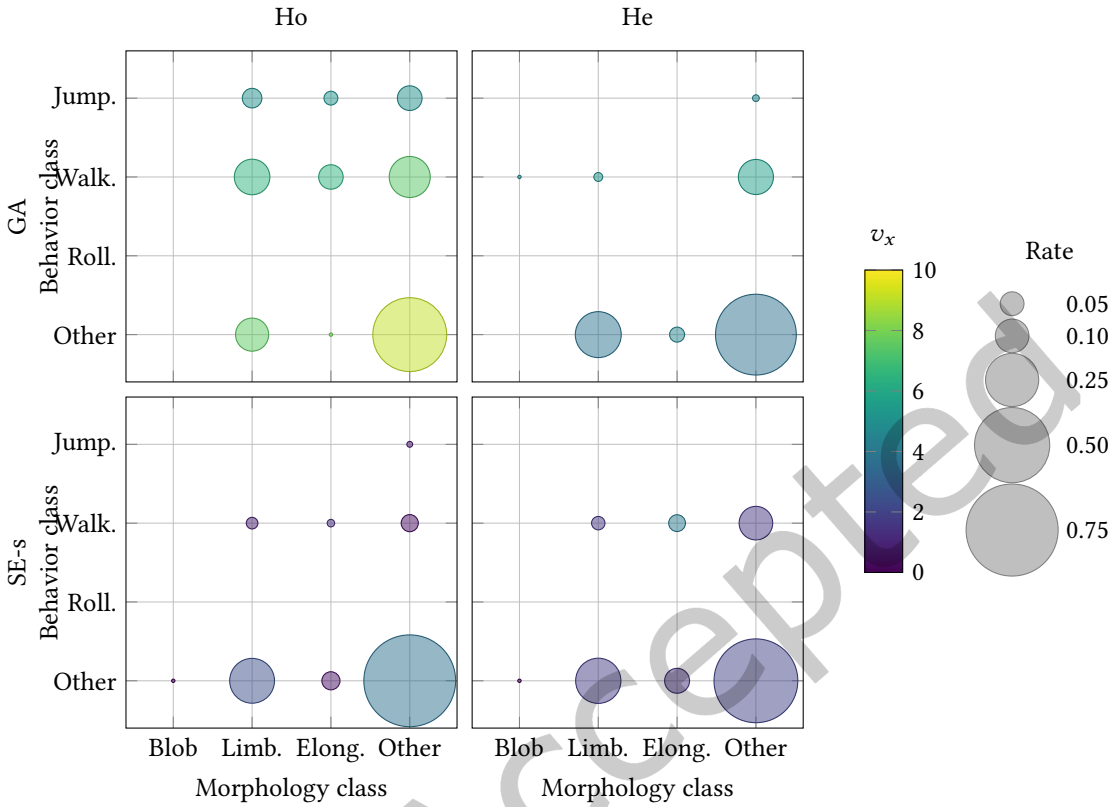


Fig. 6. Rate (bubble size) of VSRs that belong to each species, resulting from the combination of a morphology class and a behavior class, and average velocity v_x (bubble color) of the best VSR of each species, at the last generation. The plot shows median values computed across 10 evolutionary runs performed with two EAs and two controller representations.

are ill-suited for their bodies, or vice versa. To tackle this problem, other studies employed age protection [Cheney et al. 2018] or proposed to embed a learning loop within evolution [Eiben and Hart 2020; Gupta et al. 2021]. We chose to not consider these further design axes in our investigation and leave them as options for future work.

We asked ourselves whether the class differences in Figure 8 could be due to the morphology representations biasing the search towards different regions of the space of morphologies. To verify this hypothesis, we plot in Figure 9 the three morphology descriptors introduced in Section 3.3, namely elongation d_{elong} (on the x-axis), compactness d_{compact} (on the y-axis), and number of voxels d_{num} (by means of marker size) for the best individuals at the last generation. We use the descriptors computed on the "static" morphology, rather than the features $\boldsymbol{f}_{\text{morph}}$ computed out of a simulation, for two reasons: first, the descriptors are fewer; second, they are not influenced by the behavior, and are hence better suited for discussing the bias of the representation alone. We also show in Figure 10 a few sample morphologies found throughout the phylogenetic tree for Direct and GMM representations.

Figure 9 illustrates that the Direct representation covers more effectively the compactness axis, while the GMM representation covers more effectively the elongation axis. Direct best individuals are usually elongated and small in size, while GMM best individuals are all very compact and usually bigger in size. Figure 10 corroborates these observations, by showing how individuals evolved with the Direct morphology representation are, generally,

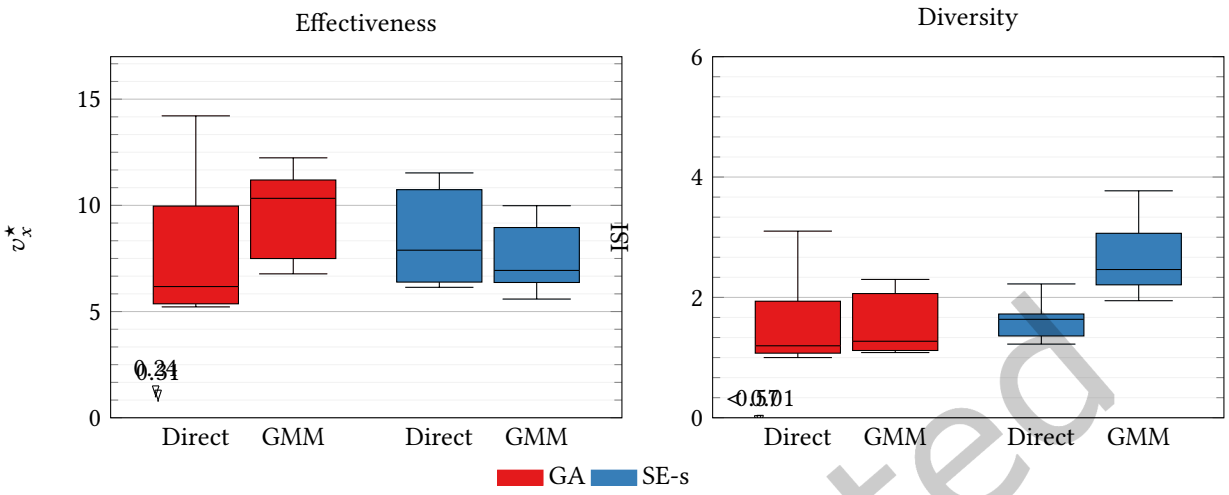


Fig. 7. Boxplots of $v_x^\star$ and ISI of the population at the last generation, obtained with four combinations of EA and morphology representation (10 evolutionary runs for each combination). Numbers above each pair of boxes are p-values. The Direct and the GMM representations are comparable in terms of effectiveness, while SE-s fosters diversity more with the GMM representation.

more knotty and less regular, with many protrusions that could potentially be employed as limbs for locomotion. This finding is particularly insightful if we consider that the Direct representation evolves very effectively in Walking individuals. Intuitively, limbs are necessary to generate a walking gait. GMM morphologies, on the other side, are more rounded and compact, sometimes spindly.

These observations raise an interesting point: the morphology representation biases the search towards particular regions of the behavior space (i.e., gaits). This conjecture is in line with the embodied cognition paradigm [Pfeifer and Bongard 2006], which posits that there is an inextricable relationship between the body and the brain. As our results point out, the Direct representation is better suited for evolving "limbed" individuals, having body extrusions that contribute to locomotion by a walking gait.

Finally, we investigated whether the biases observed so far are due to the representation, or also to the concurrent evolutionary factors of selection and genetic operators [Aaron et al. 2022]. To do so, we plot the distribution for the three morphology descriptors, d_{num}, d_{elong}, and d_{compact}, right after population initialization, and compare it with the distribution at the last generation. If a morphology representation did not introduce any bias, we would expect a uniform distribution after population initialization. Figure 11 presents the results.

For a given descriptor, selective pressure biases the search in the same direction regardless of the representation, i.e., toward more compact, elongated, and smaller morphologies, as bigger individuals are probably more difficult to evolve.

4.2 Impact of the EA

We aim at investigating what impact the EA has on effectiveness and diversity. As discussed in Section 4.1.1, we adopt Ho as controller representation. Since, as reported in Section 4.1.2, Direct and GMM representations delivered comparable results, we resort just to Direct as morphology representation for the sake of conciseness. With these settings, we performed an experimental campaign of 10 evolutionary runs with the six EAs described in Section 3.6.2, namely ES, GA, SE-g, SE-s, SE-b, and ME.

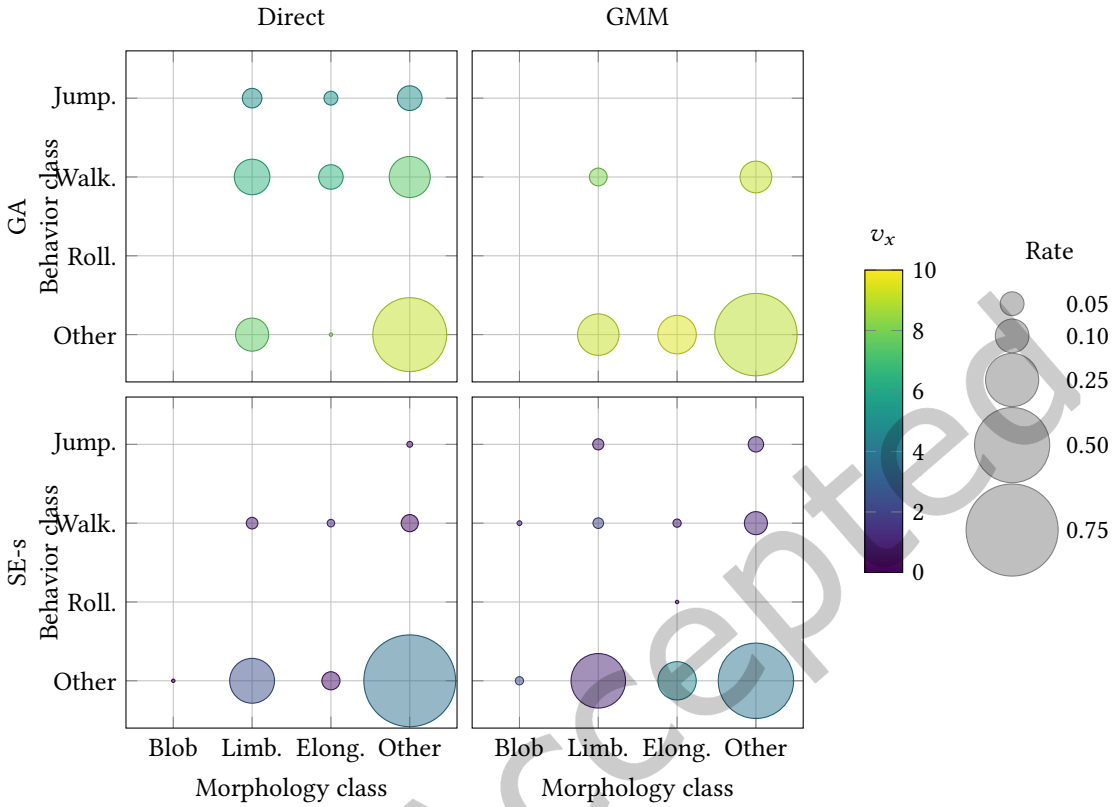


Fig. 8. Rate (bubble size) of VSRs that belong to each species, resulting from the combination of a morphology class and a behavior class, and average velocity v_x (bubble color) of the best VSR of each species, at the last generation. The plot shows median values computed across 10 evolutionary runs performed with two EAs and two morphology representations.

We show in Figure 12 how $v_x^\star$ and ISI vary over the course of evolution. We see that all EAs are able to evolve effective VSRs, i.e., they score well in terms of $v_x^\star$. In terms of ISI, SE variants and ME achieve the best results among the EAs.

Concerning diversity, the ISI plot in Figure 12 highlights some differences among the EAs: however, the variability of ISI across runs, as shown in Figure 13, is rather large for the majority of the EAs. The latter figure shows the distribution of ISI at the last generation, for all six EAs. By looking at Figure 12 it can be seen that, SE variants and ME generally maintain a large amount of diversity in the population, in terms of median value during the evolution, whereas ES and GA do not. For ES, the finding is not surprising: as discussed in Section 3.6.2, this EA does not evolve a population of actually different individuals but rather evolves one prototype individual by sampling its variants. GA performs similarly, witnessing a drastic drop in ISI around 5000 fitness evaluations. We looked at the raw results and found that the drastic drop in ISI is the outcome of the joint action of the generational model (that employs overlapping) and the crossover operator: a good individual often mates with a slightly modified copy of itself generating a "duplicate", rapidly swamping the population. We believe our GA turned out to operate with the wrong *exploration-exploitation trade-off*, a long-standing issue in EC [Črepinšek et al. 2013]. We speculate that this limitation might be addressed by employing some diversity promotion mechanism [Squillero and Tonda 2018], possibly acting at different levels of the representation (genotype, phenotype, fitness)

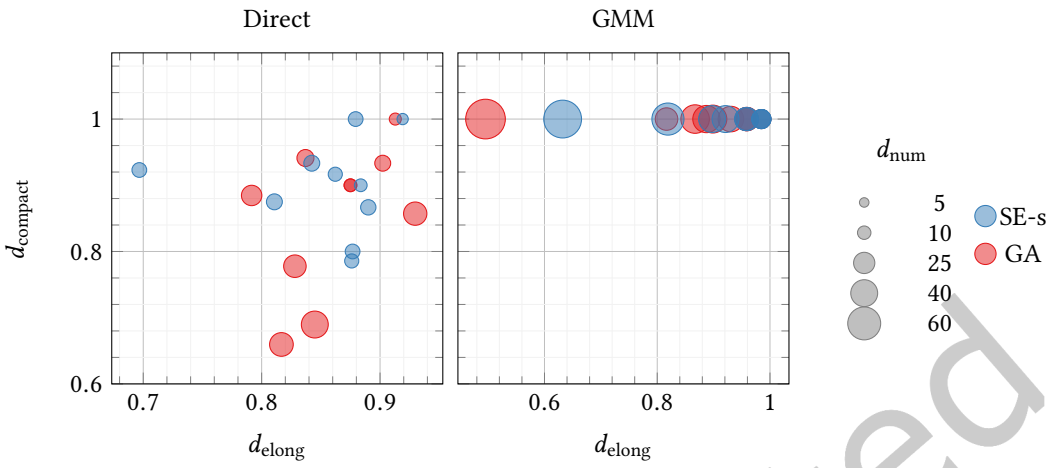


Fig. 9. Scatter plot for elongation d_{elong} (x-axis), compactness d_{compact} (y-axis), and number of voxels d_{num} (marker size) descriptors computed on the best individual of each run (GA and SE-S together), and colored by the morphology representation. Best individuals evolved with the Direct representation cover more effectively the compactness axis, while best individuals evolved with the GMM representation cover more effectively the elongation axis.

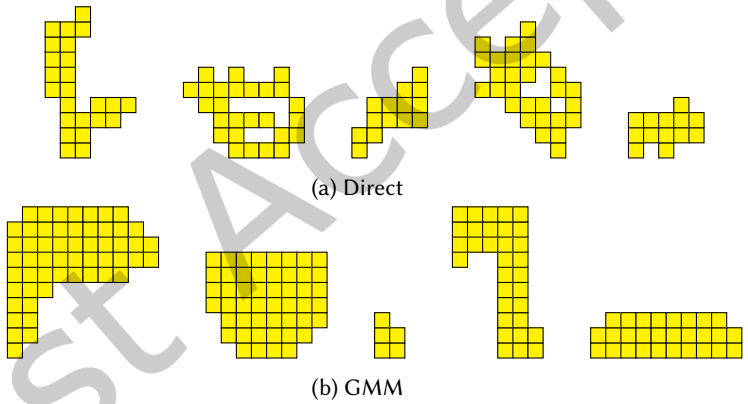


Fig. 10. Sample morphologies evolved with the Direct and GMM morphology representations. Direct individuals tend to be more "limbed", while GMM individuals tend to be more compact.

[Bartoli et al. 2019]. As expected, ME is comparable to SE variants at promoting diversity, whereas its slightly lower effectiveness may be due to its incentive for exploration: filling more cells in the archive does not necessarily entail better individuals.

Figures 14 and 15 present the breakdown by classes with the same visual syntax of Figure 6: Figure 14 shows one single plot for all the EAs; Figure 15 show six plots, one for each EA. In terms of effectiveness (color of each bubble in the figures), for the majority of the EAs, the fastest VSRs are those of class Other/Other, followed by Other/Walking; for ES, the best VSR belongs to the Limbed/Jumping class. By looking at the relative sizes of the bubbles, it can be seen that some EAs tend to favor a more even distribution of evolved VSRs across classes. Overall, Figure 15 is consistent with Figure 13: SE-s exhibits the most uneven distribution of bubble sizes. The relative majority of the evolved individuals belong to the Other behavior class, both globally and for each EA.

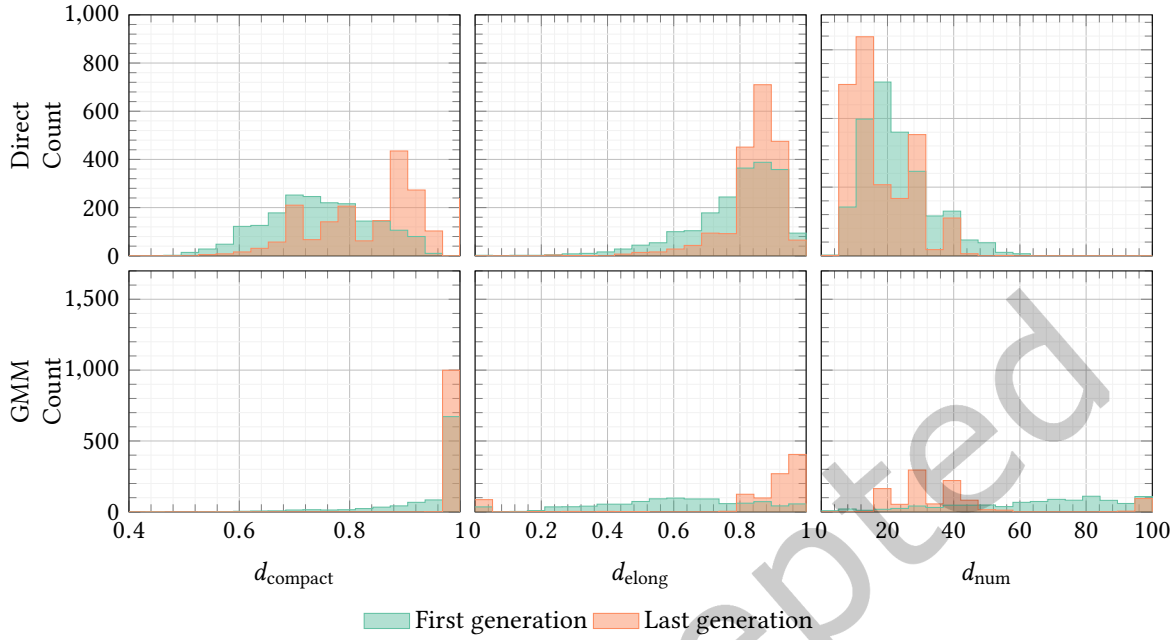


Fig. 11. Histograms for the distribution of the three morphology descriptors d_{num}, d_{elong}, and d_{compact} computed on the population of the first and last generations (of both GA and SE-s), obtained with the two morphology representations. For a given descriptor, selective pressure biases the search in the same direction regardless of the representation.

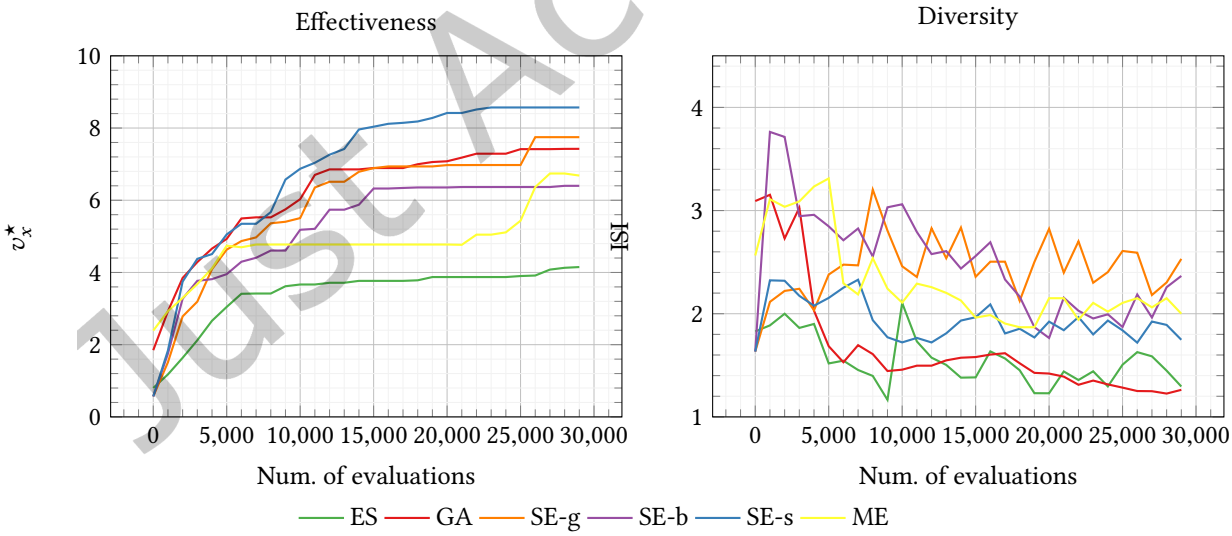


Fig. 12. $v_x^\star$ and ISI during the evolution (median values across 10 evolutionary runs for each EA). Variants of SE and ME foster diversity the most.

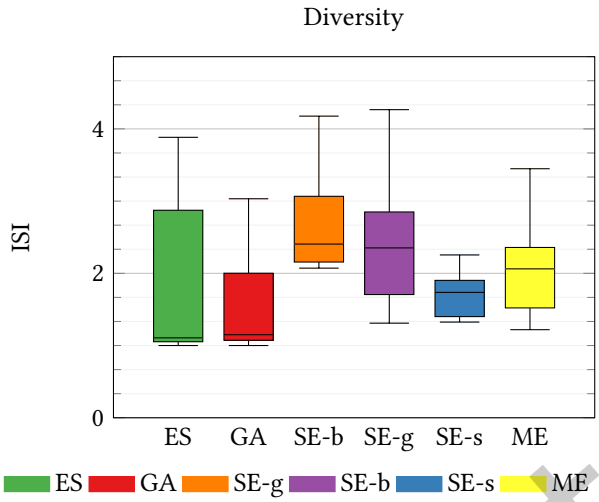


Fig. 13. Boxplots of ISI at the last generation, obtained with six EAs, Ho controller representation and Direct morphology representation (10 evolutionary runs for each EA). ES and GA exhibit the lowest median diversity.

We traced individuals of the Other behavior class to two major sub-classes: idle and vibrating. Not surprisingly, idle individuals abound, as already discussed in Section 4.1.2. Vibrating individuals manifest a behavior similar to that of Walking, but move their body at a higher frequency that makes it hard to discern which parts of the body are touching the ground and which ones are not (as a matter of example, consider Figure 2d). In this way, they achieve very high fitness. Nevertheless, the results concerning vibrating individuals call for some further remarks. If we attempted to physically realize those vibrating VSRs, maybe using the approaches of [Kriegman et al. 2020; Legrand et al. 2023; Sui et al. 2020], they would likely not be as fast as in simulation—i.e., there would likely be a *reality gap* problem [Mouret and Chatzilygeroudis 2017]. We think that the vibrating behavior evolves frequently for two reasons. First, we do not consider energy consumption in our simulations, and hence inefficient behaviors are not discouraged [Joachimczak et al. 2016]. Second, the recurrent nature of our neural controller, and in particular the voxel-to-voxel message passing, likely favors the emergence of high-frequency dynamics [Medvet et al. 2020a].

One last comment concerns the way we evaluate effectiveness. So far, we reported results in terms of $v_x^\star$, i.e., the average velocity of the best individual. In Figure 16 we plot the distribution of $\overline{v}_x$, i.e., median average velocity within the entire population of last generation, for all six EAs, side-by-side with the corresponding $v_x^\star$.

As expected, GA witnesses a non-significant difference between the best and median. Figure 16 corroborates the finding that SE variants, especially SE-g and SE-s, foster a greater amount of diversity: differences between best and median are strongly significant, attesting that speciation promotes diversity by protecting low-performing species in the population. Interestingly, with SE-b the same does not happen: from this point of view, this EA seems the one that better fits the needs of an autonomous robotic ecosystem, by preserving different but still effective species.

4.3 Impact of the environment

The third and last factor we consider in our study is the environment. In our experimental setup, the terrain profile plays the role of the environment: more arduous terrains (e.g., uphill) correspond to less hospitable environments and less arduous terrains (e.g., downhill) correspond to more hospitable environments. Based on the results of Section 4.1, we adopt Ho and Direct as controller and morphology representations, respectively, while based

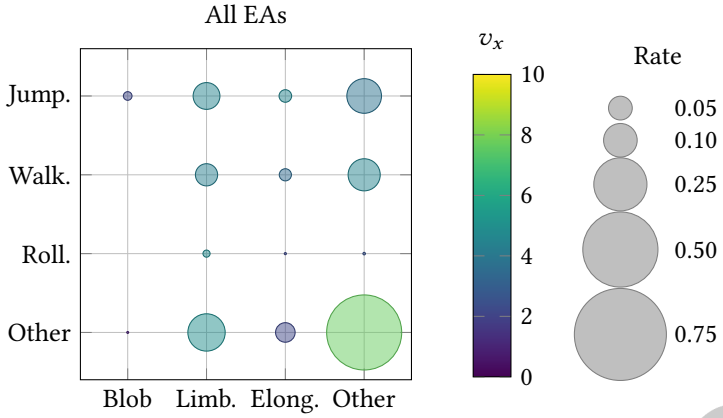


Fig. 14. Rate (bubble size) of VSRs that belong to each species, resulting from the combination of a morphology class and a behavior class, and average velocity v_x (bubble color) of the best VSR of each species, at the last generation. The plot shows median values computed across all the evolutionary runs performed with all the EAs, considered together.

on Section 4.2, we use GA and SE-s as EAs. With these settings, we performed an experimental campaign of 10 evolutionary runs with three different terrains:

(a) flat;
(b) downhill, consisting of a flat surface tilted downward by 30°;
(c) uphill, consisting of a flat surface tilted upward by 20°.

We report the results in Figure 17 in terms of $v_x^\star$ and ISI at the last generation, together with the p-values.

From the plot, we draw the following observations:

(a) downhill is the terrain favoring effectiveness the most, uphill the least;
(b) downhill appears to favor more diversity than the other two terrains, but only when using SE-s as EA;
(c) as already discussed in Section 4.2, the EA impacts diversity a lot, while the same is not true for effectiveness.

We found noticeable differences between the two new terrains in terms of the evolved species. We summarize the results in Figure 18, using the same visual syntax of Figure 6. The most effective individuals who evolved on downhill relied, for the vast majority, on Rolling to achieve very high effectiveness. To our surprise, Rolling did not amount to merely falling down the surface. Instead, successful Rolling individuals nudged themselves to create momentum at the very onset of the simulation and expanded their bodies at every roll to accompany the descent. On the other side, the most effective individuals who evolved on uphill mostly relied on crawling to clinch to the upright surface, resist gravity, and slowly move forward (a behavior labeled as Walking). Being an arduous environment, uphill generated a disproportionate amount of idle individuals of class Other, whom gravity dragged down.

As far as the morphology is concerned, downhill did favor Blob individuals, since Blob is a very suitable morphology for the Rolling behavior. On the other hand, uphill exerted high selective pressure on Blob individuals, since such morphology is not suitable for climbing an inclined surface. Instead, evolution favored Elongated individuals, who could stretch over the inclined surface and clinch to it.

At a high level, it might look like populations evolved with downhill become swamped by Rolling/Blob individuals, and so downhill does not favor diversity. This intuition is in contrast to what Figure 17 reports, as SE-s achieves a high ISI on downhill. We investigated this contradiction and found that although diverse species did indeed coexist within the same generation, Blob/Rolling tended to be the species with better effectiveness.

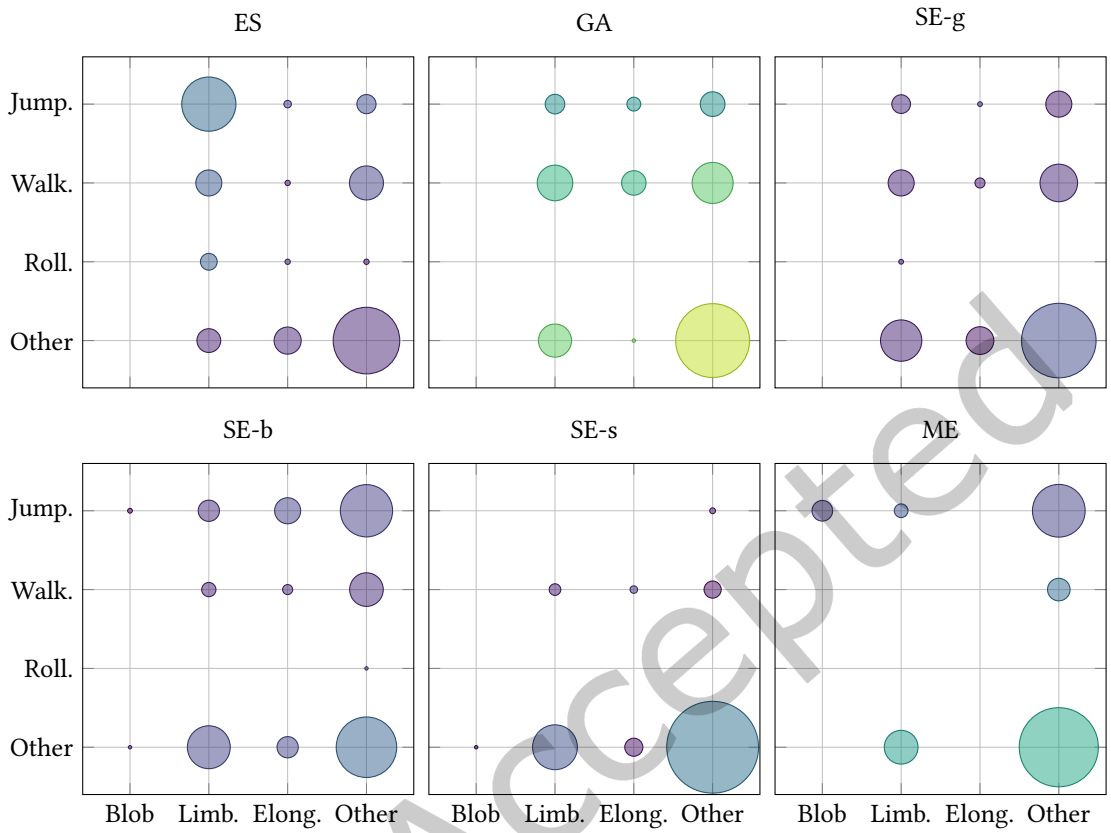


Fig. 15. Rate (bubble size) of VSRs that belong to each species, resulting from the combination of a morphology class and a behavior class, and average velocity v_x (bubble color) of the best VSR of each species, at the last generation. Each plot shows median values computed across all the evolutionary runs performed with each one of the EAs.

Finally, we considered the pool of individuals from the last generations of both EAs, and performed PCA on either their morphology features or their behavior features from Sections 3.4.1 and 3.4.2 and projected the best individual of each run on the resulting 2-D space. Figure 19 provides the results, one marker for each best individual, separately for morphology and behavior. Marker style stands for the terrain, marker color stands for morphology or behavior class.

The PCA analysis corroborates the previous discussion, as the best individuals evolved on downhill cluster together in the morphology and behavior spaces while sharing the same Blob/Rolling class. Considerations made so far for the uphill terrain stand as well.

To conclude, the effect of the environment on effectiveness reveals as expected: more hospitable environments correspond to higher values since effectiveness is an absolute measure. At the same time, diversity is higher on downhill terrain, which is a very hospitable environment, and only if the EA is capable of favoring it; the reason for this might be less competition.

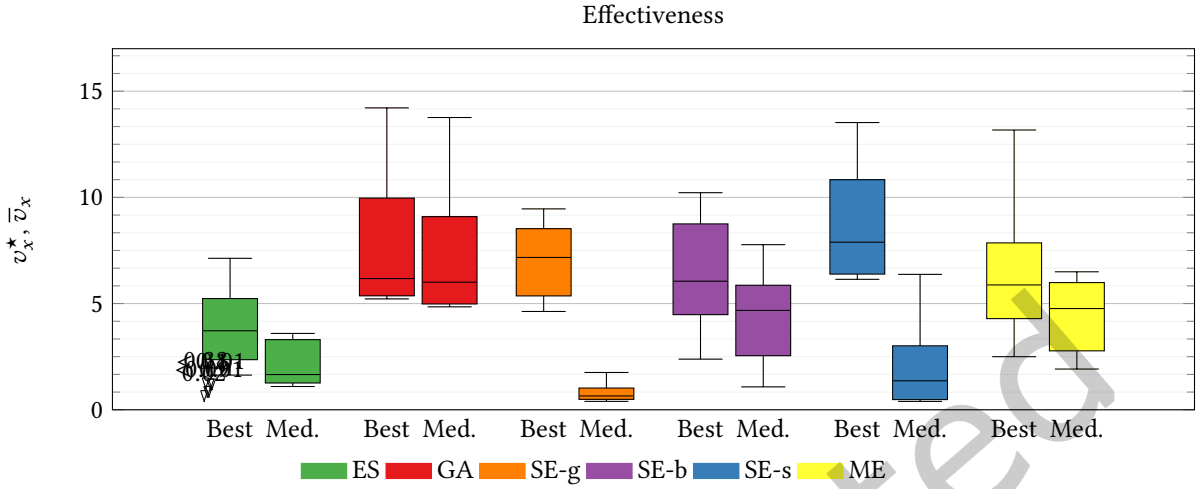


Fig. 16. Boxplots of best average velocity $v_x^\star$ and median average velocity $\overline{v}_x$ at the last generation, obtained with six EAs, Ho controller representation and Direct morphology representation (10 evolutionary runs for each EA). Numbers above each pair of boxes are p-values. SE-g and SE-s witness a dramatic difference between best and median average velocity.

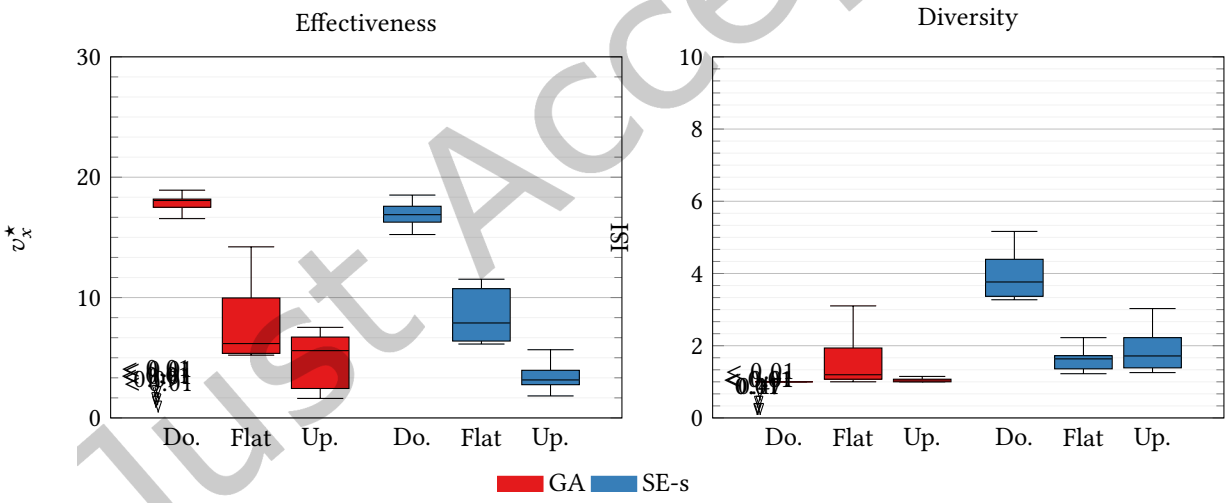


Fig. 17. Distribution of $v_x^\star$ and ISI at the last generation, obtained on three different terrains with two EAs, Ho controller representation and Direct morphology representation. Numbers above each pair of boxes are p-values. Evolution on downhill terrain finds more effective solutions, while evolution on uphill terrain finds less effective solutions. The EA impacts diversity, while it does not do so in terms of effectiveness.

5 CONCLUDING REMARKS

In this paper, we considered the automatic design of a kind of simulated modular soft robots, VSRs, by means of evolutionary optimization and investigated experimentally the impact of three key factors on the effectiveness and diversity of evolved VSRs. In the long-term vision of robotic ecosystems that are capable to stay resilient to

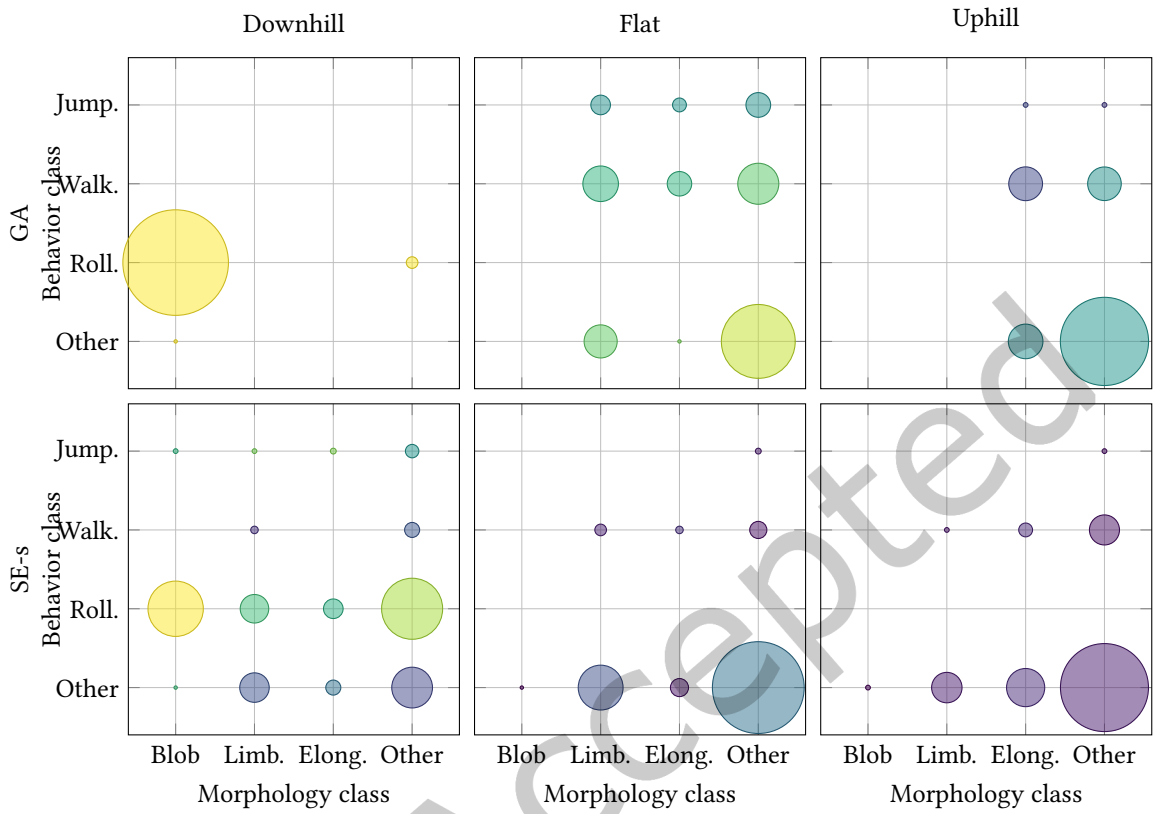


Fig. 18. Rate (bubble size) of VSRs that belong to each species, resulting from the combination of a morphology class and a behavior class, and average velocity v_x (bubble color) of the best VSR of each species, at the last generation. The plot shows median values computed across 10 evolutionary runs performed with two EAs and three environments. Bubble color and size encoding are the same of Figure 6. Downhill terrain favors Blob and Rolling individuals, while uphill favors Elongated individuals.

environmental changes even without the intervention of human designers, diversity plays a key role. However, evolving effective *and* diverse robots is not an easy task for at least two reasons. First, optimizing concurrently the morphology and the controller of effective robots is known to be difficult. Second, the interplay between quality and diversity of evolved solutions is complex in the more general context of evolutionary optimization.

We considered three key factors (representation, evolutionary algorithm, and environment) and performed several experiments in which we evolved VSRs for the task of locomotion. For analyzing the diversity of the population of evolved robots, we relied on a well-established index inspired by biology, the inverted Simpsons index, that measures how many species there are in a population and how evenly they are distributed. For assigning species to VSRs, we used a supervised machine learning approach: we defined the classes by visual inspection, collected a few examples by manual labeling, engineered suitable features, and learned a model.

We also proposed a novel EA based on a form of speciation, inspired by NEAT, and instantiated it in three variants, with species depending on the genotype, morphological traits, and behavioral traits. We showed experimentally that the proposed EA is in general able to preserve diversity without affecting effectiveness. In particular, SE-b, the variant based on behavioral traits, seems to achieve the best results. We leave to future work

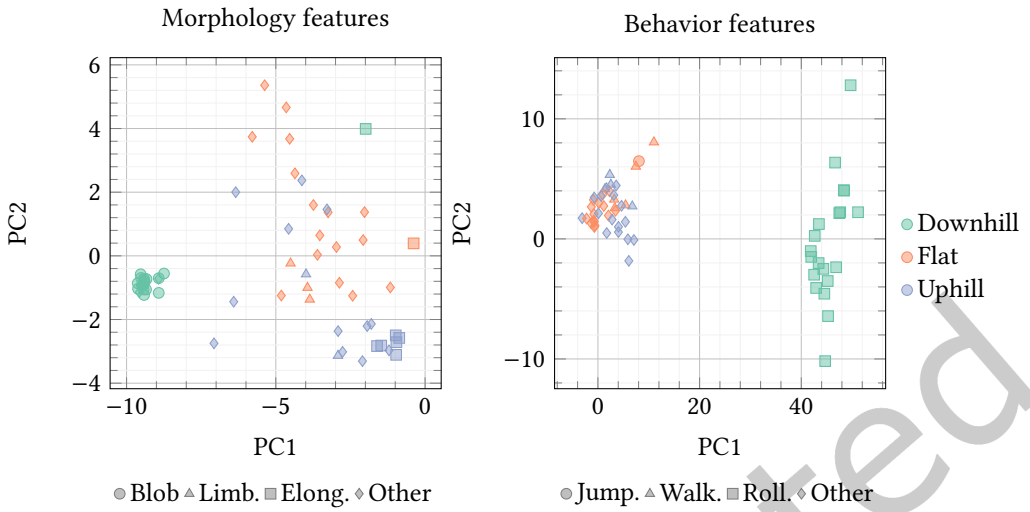


Fig. 19. PCA analysis on morphology features and behavior features. Each marker corresponds to the projection of a best individual of a run (of both GA and SE-s). Marker style stands for the terrain the individual was evolved with. Marker color stands for the morphology or behavior class. Best individuals evolved on downhill share the same morphology (Blob) and same behavior (Rolling) classes, while a majority of individuals evolved on uphill are Elongated.

the investigation on which other traits of VSRs may be useful for promoting diversity with SE-b or similar EAs. Beyond manually defined descriptors, which have the drawback of being task- and robot-specific, automatic methods like those of [Cully 2019; Paolo et al. 2020] appear promising.

Despite it is not obvious to which degree these results can be extended to other classes of robots and tasks than the one considered in this study, we believe our methodology can be easily ported to other cases. Moreover, we think that the idea of using behavioral traits for promoting diversity, as implemented in SE-b, is promising, consistently with the recent trend of quality-diversity optimization algorithms that rely on descriptors for favoring the concurrent optimization of a diverse set of solutions. Finally, we hypothesize that the diversity induced by SE-b and similar EAs might be beneficial for fighting the reality gap problem, which affects the field of automatic design of robots, not only when employing ER techniques [Salvato et al. 2021].

ACKNOWLEDGMENTS

F.P. was partially supported by a Google Faculty Research Award granted to E.M.. The experimental evaluation of this work was done on an HPC cluster within the CINECA-University of Trieste agreement. The authors wish to thank Andrea Ferigo for the MAP-Elites implementation.

REFERENCES

Eric Aaron, Joshua Hawthorne-Madell, Ken Livingston, and John H. Long. 2022. Morphological Evolution: Bioinspired Methods for Analyzing Bioinspired Robots. *Frontiers in Robotics and AI* 8 (2022).

Francesca Arese Lucini, Flaviano Morone, Maria Silvina Tomassone, and Hernán A Makse. 2020. Diversity increases the stability of ecosystems. *PloS one* 15, 4 (2020), e0228692.

Joshua E Auerbach and Josh C Bongard. 2014. Environmental influence on the evolution of morphological complexity in machines. *PLoS Comput Biol* 10, 1 (2014), e1003399.

Alberto Bartoli, Andrea De Lorenzo, Eric Medvet, and Giovanni Squillero. 2019. Multi-level diversity promotion strategies for Grammar-guided Genetic Programming. *Applied Soft Computing* 83 (2019), 105599.

Hans-Georg Beyer and Hans-Paul Schwefel. 2002. Evolution strategies–a comprehensive introduction. *Natural computing* 1, 1 (2002), 3–52.

Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
Edgar Buchanan, Léni K Le Goff, Wei Li, Emma Hart, Agoston E Eiben, Matteo De Carlo, Alan F Winfield, Matthew F Hale, Robert Woolley, Mike Angus, et al. 2020. Bootstrapping Artificial Evolution to Design Robots for Autonomous Fabrication. *Robotics* 9, 4 (2020), 106.
Wilhelm Burger, Mark James Burge, Mark James Burge, and Mark James Burge. 2009. *Principles of digital image processing*. Vol. 111. Springer.
Nick Cheney, Josh Bongard, and Hod Lipson. 2015. Evolving soft robots in tight spaces. In *Proceedings of the 2015 annual conference on Genetic and Evolutionary Computation*. ACM, 935–942.
Nick Cheney, Josh Bongard, Vytas SunSpiral, and Hod Lipson. 2018. Scalable co-optimization of morphology and control in embodied machines. *Journal of The Royal Society Interface* 15, 143 (2018), 20170937.
Nick Cheney, Robert MacCurdy, Jeff Clune, and Hod Lipson. 2013. Unshackling evolution: evolving soft robots with multiple materials and a powerful generative encoding. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*. ACM, 167–174.
Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. 2018. Back to basics: Benchmarking canonical evolution strategies for playing atari. *arXiv preprint arXiv:1802.08842* (2018).
James W Cooley and John W Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation* 19, 90 (1965), 297–301.
Matej Črepinšek, Shih-Hsi Liu, and Marjan Mernik. 2013. Exploration and exploitation in evolutionary algorithms: A survey. *ACM Computing Surveys (CSUR)* 45, 3 (2013), 1–33.
Antoine Cully. 2019. Autonomous skill discovery with quality-diversity and unsupervised descriptors. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 81–89.
Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature* 521, 7553 (2015), 503–507.
Antoine Cully and Yiannis Demiris. 2017. Quality and diversity optimization: A unifying modular framework. *IEEE Transactions on Evolutionary Computation* 22, 2 (2017), 245–259.
Matteo De Carlo, Daan Zeeuwe, Eliseo Ferrante, Gerben Meynen, Jacintha Ellers, and AE Eiben. 2020. Influences of Artificial Speciation on Morphological Robot Evolution. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2272–2279.
Kenneth A De Jong. 2006. *Evolutionary Computation: A Unified Approach*. MIT Press.
AE Eiben and Emma Hart. 2020. If it evolves it needs to learn. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*. 1383–1384.
Andrea Ferigo, Giovanni Iacca, Eric Medvet, and Federico Pigozzi. 2021. Evolving Hebbian Learning Rules in Voxel-based Soft Robots. (2021).
Andrea Ferigo, Eric Medvet, and Giovanni Iacca. 2022a. Optimizing the Sensory Apparatus of Voxel-Based Soft Robots Through Evolution and Babbling. *SN Computer Science* 3, 2 (2022), 1–17.
Andrea Ferigo, LB Soros, Eric Medvet, and Giovanni Iacca. 2022b. On the Entanglement between Evolvability and Fitness: an Experimental Study on Voxel-based Soft Robots. In *ALIFE 2022: The 2022 Conference on Artificial Life*. MIT Press.
Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research* 15, 1 (2014), 3133–3181.
Steven E Franklin and Oumer S Ahmed. 2018. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. *International Journal of Remote Sensing* 39, 15-16 (2018), 5236–5245.
Agrim Gupta, Silvio Savarese, Surya Ganguli, and Li Fei-Fei. 2021. Embodied Intelligence via Learning and Evolution. *arXiv preprint arXiv:2102.02202* (2021).
David Ha. 2019. Reinforcement learning for improving agent design. *Artificial life* 25, 4 (2019), 352–365.
M Hale, Edgar Buchanan Berumen, Alan Winfield, Jon Timmis, Emma Hart, Gusz Eiben, Wei Li, and Andy Tyrrell. 2019. The are robot fabricator: How to (re) produce robots that can evolve in the real world. In *International Society for Artificial Life: ALIFE2019*. York, 95–102.
Matthew F Hale, Mike Angus, Edgar Buchanan, Wei Li, Robert Woolley, Léni K Le Goff, Matteo De Carlo, Jon Timmis, Alan F Winfield, Emma Hart, et al. 2020. Hardware design for autonomous robot evolution. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2140–2147.
Nikolaus Hansen and Andreas Ostermeier. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proceedings of IEEE international conference on evolutionary computation*. IEEE, 312–317.
Jonathan Hiller and Hod Lipson. 2012. Automatic design and manufacture of soft robots. *IEEE Transactions on Robotics* 28, 2 (2012), 457–466.
Gregory S Hornby, Jordan B Pollack, et al. 2001. Body-brain co-evolution using L-systems as a generative encoding. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*. 868–875.
Michał Joachimczak, Reiji Suzuki, and Takaya Arita. 2016. Artificial Metamorphosis: Evolutionary Design of Transforming, Soft-Bodied Robots. *Artificial Life* 22, 3 (2016), 271–298.
Sangbae Kim, Cecilia Laschi, and Barry Trimmer. 2013. Soft robotics: a bioinspired evolution in robotics. *Trends in biotechnology* 31, 5 (2013), 287–294.
Sam Kriegman, Douglas Blackiston, Michael Levin, and Josh Bongard. 2020. A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences* 117, 4 (2020), 1853–1859.

Sam Kriegman, Douglas Blackiston, Michael Levin, and Josh Bongard. 2021. Kinematic self-replication in reconfigurable organisms. *Proceedings of the National Academy of Sciences* 118, 49 (2021).

Julie Legrand, Seppe Terryn, Ellen Roels, and Bram Vanderborght. 2023. Reconfigurable, multi-material, voxel-based soft robots. *IEEE Robotics and Automation Letters* (2023).

Joel Lehman and Kenneth O Stanley. 2008. Exploiting Open-Endedness to Solve Problems Through the Search for Novelty. *Artificial Life* 11, 329.

Joel Lehman and Kenneth O Stanley. 2011. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*. 211–218.

Bruce G Lindsay. 1995. Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics*. JSTOR, i–163.

Hod Lipson, Vytas Sunspiral, Josh Bongard, and Nicholas Cheney. 2016. On the difficulty of co-optimizing morphology and control in evolved virtual creatures. In *Artificial Life Conference Proceedings 13*. MIT Press, 226–233.

Stuart Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (1982), 129–137.

Anne E Magurran. 2013. *Measuring biological diversity*. John Wiley & Sons.

Eric Medvet, Alberto Bartoli, Andrea De Lorenzo, and Giulio Fidel. 2020a. Evolution of distributed neural controllers for voxel-based soft robots. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. 112–120.

Eric Medvet, Alberto Bartoli, Andrea De Lorenzo, and Stefano Seriani. 2020b. 2D-VSR-Sim: A simulation tool for the optimization of 2-D voxel-based soft robots. *SoftwareX* 12 (2020).

Eric Medvet, Alberto Bartoli, Andrea De Lorenzo, and Stefano Seriani. 2020c. Design, Validation, and Case Studies of 2D-VSR-Sim, an Optimization-friendly Simulator of 2-D Voxel-based Soft Robots. *arXiv* (2020), arXiv–2001.

Eric Medvet, Alberto Bartoli, Federico Pigozzi, and Marco Rochelli. 2021. Biodiversity in evolved voxel-based soft robots. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 129–137.

Karine Miras, Eliseo Ferrante, and AE Eiben. 2020. Environmental influences on evolvable robots. *PloS one* 15, 5 (2020), e0233848.

Jean-Baptiste Mouret and Konstantinos Chatzilygeroudis. 2017. 20 years of reality gap: a few thoughts about simulators in evolutionary robotics. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1121–1124.

J-B Mouret and Stéphane Doncieux. 2012. Encouraging behavioral diversity in evolutionary robotics: An empirical study. *Evolutionary computation* 20, 1 (2012), 91–133.

Giorgia Nadizar, Eric Medvet, Felice Andrea Pellegrino, Marco Zullich, and Stefano Nichele. 2021. On the effects of pruning on evolved neural controllers for soft robots. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1744–1752.

Geoff Nitschke and David Howard. 2021. AutoFac: The Perpetual Robot Machine. *IEEE Transactions on Artificial Intelligence* (2021).

Stefano Nolfi. 2021. *Behavioral and Cognitive Robotics: An adaptive perspective*. Stefano Nolfi.

Stefano Nolfi and Dario Floreano. 2000. *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines*. MIT press.

Jørgen Nordmoen, Frank Veenstra, Kai Olav Ellefsen, and Kyrre Glette. 2021. MAP-Elites enables Powerful Stepping Stones and Diversity for Modular Robotics. *Frontiers in Robotics and AI* 8 (2021).

Arthur O'Sullivan and Steven M Sheffrin. 2003. Economics: Principles in action. Upper Saddle River, New Jersey 07458: Pearson Prentice Hall. (2003).

Paolo Pagliuca and Stefano Nolfi. 2020. The Dynamic of Body and Brain Co-Evolution. *arXiv preprint arXiv:2011.11440* (2020).

Giuseppe Paolo, Alban Laflaquiere, Alexandre Coninx, and Stephane Doncieux. 2020. Unsupervised learning and exploration of reachable outcome space. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2379–2385.

Rolf Pfeifer and Josh Bongard. 2006. *How the body shapes the way we think: a new view of intelligence*. MIT press.

Franz Rothlauf. 2006. Representations for genetic and evolutionary algorithms. In *Representations for Genetic and Evolutionary Algorithms*. Springer, 9–32.

David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. 1986. Learning internal representations by back-propagating errors. *Nature* 323, 99 (1986), 533–536.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. 2017. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864* (2017).

Erica Salvato, Gianfranco Fenu, Eric Medvet, and Felice Andrea Pellegrino. 2021. Crossing the Reality Gap: a Survey on Sim-to-Real Transferability of Robot Controllers in Reinforcement Learning. *IEEE Access* (2021).

Eivind Samuelsen and Kyrre Glette. 2014. Some distance measures for morphological diversification in generative evolutionary robotics. In *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*. 721–728.

Franziska Schrodt, Joseph J Bailey, W Daniel Kissling, Kenneth F Rijsdijk, Arie C Seijmonsbergen, Derk Van Ree, Jan Hjort, Russell S Lawley, Christopher N Williams, Mark G Anderson, et al. 2019. Opinion: To advance sustainable stewardship, we must document not only biodiversity but geodiversity. *Proceedings of the National Academy of Sciences* 116, 33 (2019), 16155–16158.

Hans-Paul Schwefel. 1965. Cybernetic evolution as strategy for experimental research in fluid mechanics. *Master's Thesis, Technical University of Berlin* (1965).

Fernando Silva, Miguel Duarte, Luís Correia, Sancho Moura Oliveira, and Anders Lyhne Christensen. 2016. Open issues in evolutionary robotics. *Evolutionary computation* 24, 2 (2016), 205–236.

Edward H Simpson. 1949. Measurement of diversity. *Nature* 163, 4148 (1949), 688–688.

Karl Sims. 1994. Evolving virtual creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, 15–22.

Giovanni Squillero and Alberto Tonda. 2016. Divergence of character and premature convergence: A survey of methodologies for promoting diversity in evolutionary optimization. *Information Sciences* 329 (2016), 782–799.

Giovanni Squillero and Alberto Tonda. 2018. Promoting diversity in evolutionary optimization: why and how. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 998–1016.

Kenneth O Stanley and Risto Miikkulainen. 2002. Evolving neural networks through augmenting topologies. *Evolutionary computation* 10, 2 (2002), 99–127.

Xin Sui, Hegao Cai, Dongyang Bie, Yu Zhang, Jie Zhao, and Yanhe Zhu. 2020. Automatic generation of locomotion patterns for soft modular reconfigurable robots. *Applied Sciences* 10, 1 (2020), 294.

Michael A Tabak, Mohammad S Norouzzadeh, David W Wolfson, Steven J Sweeney, Kurt C VerCauteren, Nathan P Snow, Joseph M Halseth, Paul A Di Salvo, Jesse S Lewis, Michael D White, et al. 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution* 10, 4 (2019), 585–590.

Jacopo Talamini, Eric Medvet, Alberto Bartoli, and Andrea De Lorenzo. 2019. Evolutionary Synthesis of Sensing Controllers for Voxel-based Soft Robots. In *Artificial Life Conference Proceedings*. MIT Press, 574–581.

Jacopo Talamini, Eric Medvet, and Stefano Nichele. 2021. Criticality-Driven Evolution of Adaptable Morphologies of Voxel-Based Soft-Robots. *Frontiers in Robotics and AI* 8 (2021), 172.

Danesh Tarapore, Jeff Clune, Antoine Cully, and Jean-Baptiste Mouret. 2016. How do different encodings influence the performance of the MAP-Elites algorithm?. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. 173–180.

David Tilman, Michael Clark, David R Williams, Kaitlin Kimmel, Stephen Polasky, and Craig Packer. 2017. Future threats to biodiversity and pathways to their prevention. *Nature* 546, 7656 (2017), 73–81.

Leonardo Trujillo, Gustavo Olague, Evelyne Lutton, Francisco Fernandez de Vega, León Dozal, and Eddie Clemente. 2011. Speciation in behavioral space for evolutionary robotics. *Journal of Intelligent & Robotic Systems* 64, 3 (2011), 323–351.

Vassiiis Vassiliades and Jean-Baptiste Mouret. 2018. Discovering the elite hypervolume by leveraging interspecies correlation. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 149–156.

Frank Veenstra, Andres Faina, Sebastian Risi, and Kasper Stoy. 2017. Evolution and morphogenesis of simulated modular robots: a comparison between a direct and generative encoding. In *European Conference on the Applications of Evolutionary Computation*. Springer, 870–885.

Michael Wainberg, Babak Alipanahi, and Brendan J Frey. 2016. Are random forests truly the best classifiers? *The Journal of Machine Learning Research* 17, 1 (2016), 3837–3841.

Mark Yim, Wei-Min Shen, Behnam Salemi, Daniela Rus, Mark Moll, Hod Lipson, Eric Klavins, and Gregory S Chirikjian. 2007. Modular self-reconfigurable robot systems [grand challenges of robotics]. *IEEE Robotics & Automation Magazine* 14, 1 (2007), 43–52.

Crawford Young. 1979. *The politics of cultural pluralism*. Univ of Wisconsin Press.