

Discovering Attention-Based Genetic Algorithms via Meta-Black-Box Optimization

Robert Tjarko Lange*
Technical University Berlin

Tom Schaul, Yutian Chen
DeepMind

Chris Lu*†, Tom Zahavy‡
†University of Oxford, ‡DeepMind

Valentin Dalibard, Sebastian Flennerhag
DeepMind

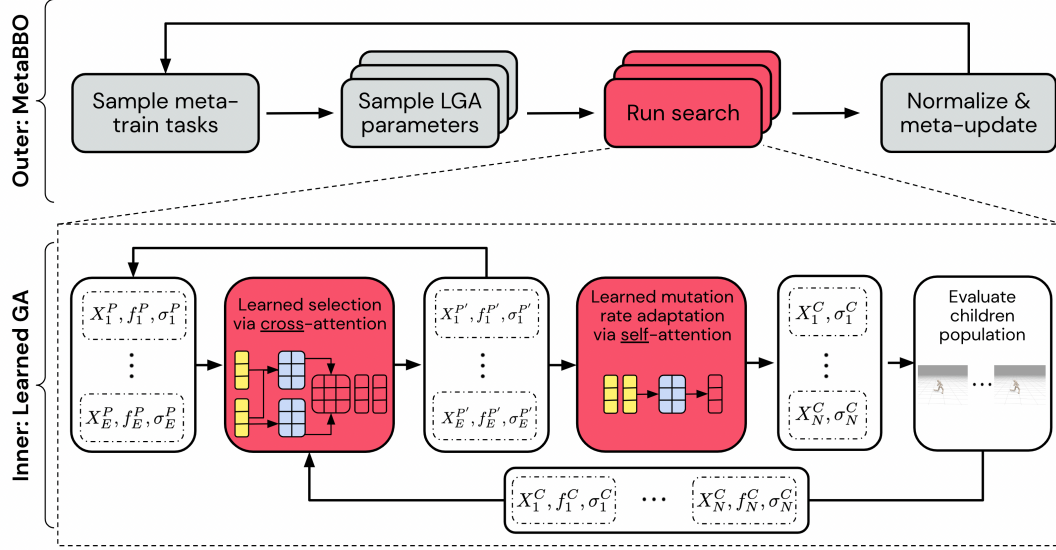


Figure 1: Discovering attention-based Learned Genetic Algorithms (LGA) via MetaBBO. At each meta-iteration one samples a set of inner loop tasks and a set of candidate LGA parameters from a meta-evolutionary optimizers (EO). Afterwards, one runs an inner loop search and compute a normalized meta-fitness score across the tasks. We update the meta-EO and iterate.

ABSTRACT

Genetic algorithms constitute a family of black-box optimization algorithms, which take inspiration from the principles of biological evolution. While they provide a general-purpose tool for optimization, their particular instantiations can be heuristic and motivated by loose biological intuition. In this work we explore a fundamentally different approach: Given a sufficiently flexible parametrization of the genetic operators, we discover entirely new genetic algorithms in a data-driven fashion. More specifically, we parametrize selection and mutation rate adaptation as cross- and self-attention modules and use *Meta-Black-Box-Optimization* to evolve their parameters on a set of diverse optimization tasks. The resulting *Learned Genetic Algorithm* outperforms state-of-the-art adaptive

baseline genetic algorithms and generalizes far beyond its meta-training settings. The learned algorithm can be applied to previously unseen optimization problems, search dimensions & evaluation budgets. We conduct extensive analysis of the discovered operators and provide ablation experiments, which highlight the benefits of flexible module parametrization and the ability to transfer (‘plug-in’) the learned operators to conventional genetic algorithms.

CCS CONCEPTS

• Computing methodologies → Genetic algorithms;

KEYWORDS

genetic algorithm, machine learning, meta-learning

* Work done as interns at DeepMind. Contact: robert.t.lange@tu-berlin.de.

ACM Reference Format:

Robert Tjarko Lange*, Tom Schaul, Yutian Chen, Chris Lu*†, Tom Zahavy‡, and Valentin Dalibard, Sebastian Flennerhag. 2023. Discovering Attention-Based Genetic Algorithms via Meta-Black-Box Optimization. In *Genetic and Evolutionary Computation Conference (GECCO '23)*, July 15–19, 2023, Lisbon, Portugal. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3583131.3590496>

1 INTRODUCTION

Motivation. Genetic algorithms (GAs) provide a set of evolution-inspired optimization algorithms, which are flexibly applicable to black-box optimization (BBO) problems. They commonly rely on human designed operators, which impose a restrictive and most importantly *subjective* set of manual priors. This bears the risk of domain overfitting and limited generalization capabilities. Based on recent results in the discovery of attention-based Evolution Strategies [25], we propose that these limitations can be overcome by meta-learning effective GA operators from data. Thereby, the inductive biases of the GA itself are indirectly encoded by its parametrization & can be discovered in an optimization-driven fashion, i.e. by improving its meta-performance on a distribution of relevant tasks.

Approach. Inspired by the recent success of the Set Transformer [27] architecture, we introduce neural network-based architectures to substitute core genetic operators: *Selection* and *mutation rate adaptation* are cast as dot-product attention modules, which can flexibly be applied to problems with varying dimensions & population sizes. The resulting family of genetic algorithms can implement different operations based on the specific module weights. We meta-evolve these weights on a set of representative optimization problems using meta-black-box optimization (*MetaBBO*, [25]).

Results. We evaluate the performance and generalization capabilities of the meta-trained Learned Genetic Algorithm (*LGA*). LGA is capable of generalizing far beyond its meta-training distribution and outperforms established baseline GAs on several BBO benchmarks (BBOB) [2, 13] and neuroevolution problems. This includes different optimization functions, number of search dimensions and population sizes. Furthermore, we analyze the discovered neural network GA operators: The selection operator has learned an adaptive form of truncation, which maintains diversity and redundancy among the parent population. The learned mutation rate adaption operator, on the other hand, automatically scales the amount of exploration in a task-dependent fashion. We investigate the importance of the meta-evolution task distribution: While it is possible to meta-evolve effective LGAs on as little as five BBOB functions, we show that LGAs can overfit their meta-training distribution and one has to ensure sufficient meta-regularization for broad generalization. Trained LGAs are robust to their choice of hyperparameters and the details of their meta-training procedure. Finally, we show that the individual learned operators can replace white-box GA operators inducing a positive transfer effect.

Contributions. Our contributions are summarized as follows:

- (1) We propose a dot-product attention-based parametrization of the selection & mutation rate adaptation GA operators.
- (2) We discover new GAs by meta-evolving their parameters based on the performance on meta-training BBO tasks.
- (3) The resulting LGA is capable of generalizing far beyond its meta-training settings and outperforms several GA baselines on various benchmark tasks and evaluation budgets.
- (4) We perform several ablations to LGA to assess the contributions of learning the individual operator modules. Both learned operators contribute to the overall performance.
- (5) We highlight the robustness of the trained LGA to its hyperparameters, the transferability of the learned components and the stability of the MetaBBO procedure.

2 RELATED WORK

Discovery via Meta-Learned Algorithms. Recent efforts have proposed to replace manually designed algorithms by end-to-end optimized inductive biases, by meta-learning parametrized components on a representative task distribution. E.g. this includes the discovery of Reinforcement Learning objectives [28, 33, 44], schedules of algorithm hyperparameters via meta-gradients [9, 34, 45, 47], and the meta-learning of entire learning algorithms [18, 19, 42]. The discovery process can be supported by suitable neural network architectures. Our proposed LGA architecture leverages attention layers to derive a neural network-based family of GAs.

Meta-Learned Gradient-Based Optimization. Our work is closely related to the ambition of meta-learning gradient descent-based learning rules [1, 3, 31]. These approaches rely on access to efficient gradient calculations via the backpropagation algorithm and thereby do not apply to BBO problems. A small neural network processes the gradient and standard optimizer statistics (momentum, etc.) to output a weight change. The optimizer network weights in turn have been meta-learned on a task distribution [30]. Metz et al. [29] showed that this results in a highly competitive optimizer for deep learning tasks. Our MetaBBO-discovered LGA, on the other hand, provides a general-purpose BBO, which does not require differentiable objective functions.

Meta-Learned Population-Based Optimization. Shala et al. [38] meta-learn a controller for the scalar mutation rate in CMA-ES [15]. Chen et al. [5], Gomes et al. [12], TV et al. [41] previously optimized entire neural network-parametrized algorithms for low-dimensional BBO. All of them use a recurrent network, which processes raw solution candidates and their respective fitness scores. These methods often struggle to generalize to new optimization domains and are often constrained to fixed population sizes and/or search dimensions. Lange et al. [25] recently leveraged the equivariance property of dot-product self-attention to the input ordering [20, 27, 39] to learn adaptive recombination weights for evolution strategies. The proposed LGA extends this attention-based BBO perspective in order to characterize GA operators. After successful meta-training, the learned GA is capable of generalizing to unseen populations and large search spaces. To the best of our knowledge we are the first to demonstrate that a meta-learned GA generalizes to challenging neuroevolution tasks. Finally, the MetaBBO approach does not require access to knowledge of meta-task optima [41] or a teacher algorithm [38].

Baseline Genetic Algorithms. Throughout the paper we compare against four competitive baseline GAs including the following:

- Gaussian GA [35]: A simple GA with Gaussian perturbations and fixed mutation strength using truncation selection.
- MR-1/5 GA [35]: Doubles the mutation rate if 1/5 of all perturbations were beneficial. Otherwise, the rate is halved.
- SAMR-GA [7]: Self-adapts per-parent mutation rates based on a simple co-evolution heuristic and meta-mutation rate.
- GESMR-GA [21]: Avoids vanishing mutation rates by using a group elite selection criterion and mutation rate sharing.

We additionally consider Sep-CMA-ES [36] as a scalable evolution strategy baseline for neuroevolution tasks. Each baseline GA was tuned using small grid search sweeps (see Appendix B). Otherwise, we adopted the settings provided by the authors.

3 BACKGROUND

Black-Box Optimization. Throughout this manuscript, we are interested in efficient continuous black-box optimization: Given a function $f(x) : \mathbb{R}^D \rightarrow \mathbb{R}$ with unknown functional form, i.e. we cannot compute its derivative, we seek to find its global optimum:

$$\min_x f(x), \text{ s.t. } x_d \in [l_d, u_d] \subset [-\infty, \infty], \forall d = 1, \dots, D,$$

Genetic Algorithms. GAs provide a class of BBO algorithms, which iteratively evaluate a population consisting of N solution candidates $X^C = [x_1^C, \dots, x_N^C]^T \in \mathbb{R}^{N \times D}$ ('children') with fitness $f^C \in \mathbb{R}^N$. Given a set of E 'parent' solutions $X^P = [x_1^P, \dots, x_E^P]^T \in \mathbb{R}^{E \times D}$ with associated fitness $f^P \in \mathbb{R}^E$, the parents are replaced by the children using a heuristic fitness-based selection criterion. Most GAs make use of truncation selection in which all children and parents $[x_1^C, \dots, x_N^C, x_1^P, \dots, x_E^P]^T$ are jointly sorted by their fitness. The top- E performing solutions replace the parent archive:

$$X^{P'}, f^{P'} = \text{Selection}(X^P, f^P, X^C, f^C).$$

Commonly the number of parents is set to be small $E \ll N$ and often even $E = 1$, enforcing a type of hill climbing. N children candidates are uniformly sampled with replacement from the parents:

$$\tilde{X}^P, \tilde{f}^P = \text{Sample}(X^P, f^P) \in \mathbb{R}^{N \times D}.$$

Afterwards, they are perturbed using a mutation rate (MR) $\sigma \in \mathbb{R}_+$, which controls the strength of the Gaussian noise, $\epsilon_j \sim \mathcal{N}(0_D, I_D)$.

$$X_j^C = \text{Mutation}(\tilde{X}_j^P, \sigma) = \tilde{X}_j^P + \sigma \epsilon_j \in \mathbb{R}^D, \forall j = 1, \dots, N.$$

Many competitive GAs keep a vector of parent-specific [7] mutation rates $\sigma^P \in \mathbb{R}^E$ and additionally perform an intermediate mutation rate adaptation (MRA) step to improve the mutation rate(s) given information gathered throughout the fitness evaluations:

$$\sigma^P = \text{Adaptation}(f^P, \sigma^P, f^C, \sigma^C) \in \mathbb{R}^E.$$

In this case, the children are perturbed by their sampled individual-specific mutation rate ($\sigma^C \leftarrow \tilde{\sigma}^P \in \mathbb{R}^N$) and selection also applies to the children's MR. Alternatively, GESMR-GA [21] forms parent sub-groups and online co-evolves group-level mutation rates based on their observed fitness improvements.

Set Operations via Dot-Product Self-Attention. Scaled dot-product attention (SDPA) is especially well suited to characterize algorithms performing set operations, since it naturally enforces a permutation invariant function. Consider the standard formulation of SDPA, which embeds a set of N input tokens, $X \in \mathbb{R}^{N \times D}$, into D_K -dimensional latent query Q , key K and value V representations:

$$Q = XW_Q \in \mathbb{R}^{N \times D_K},$$

$$K = XW_K \in \mathbb{R}^{N \times D_K},$$

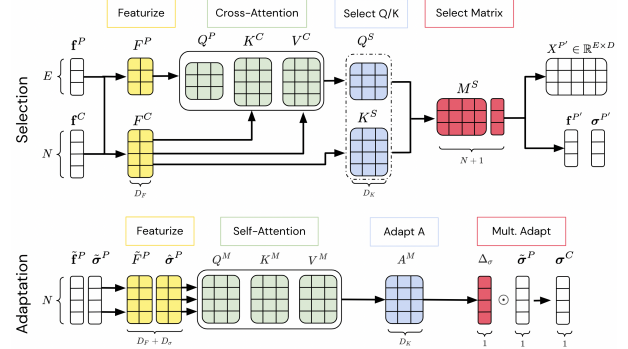
$$V = XW_V \in \mathbb{R}^{N \times D_K}.$$

The output Y is computed as a linear combination of the values:

$$Y = \text{softmax}\left(QK^T / \sqrt{D_K}\right)V \in \mathbb{R}^{N \times D_K}.$$

It can be shown that this transformation is equivariant to the ordering of the tokens in X , i.e. permuting the rows of X will apply the same permutation to the rows of Y [20, 39]. We will leverage this suitable inductive bias to characterize GAs, which inherently operate on sets of solution candidates and their fitness scores.

Figure 2: Learned Genetic Operators. Top: Cross-attention selection between parent & children fitness features. Bottom: MRA via self-attention on parent mutation & fitness features.



4 ATTENTION-BASED GENETIC OPERATORS

We now introduce an attention-based parametrization of the genetic Selection and Adaptation operators (Figure 2). These in turn will be meta-optimized on a set of representative optimization tasks in order to capture useful BBO mechanisms. We start by answering a natural question: What inputs should be processed by the operators in order to enable generalization across fitness and solution scales?

Attention Features via Fitness Scores & Mutation Strength. To compute attention scores across parents and children, we need to construct sufficient features, which modulate effective selection and mutation rate adaptation. Furthermore, we want the meta-learned operations to generalize across different test optimization scales. Hence, we consider scale invariant normalizations: E.g. z-scoring and centered ranks (in $[-0.5, 0.5]$). We transform both the raw fitness scores (D_F dim.) and the parent mutation rates (D_σ dim.) to construct a set of features processed by the attention layers:

- (1) $F \in \mathbb{R}^{(N+E) \times D_F}$: Joint fitness transformations of parents and children (z-scores & centered ranks).
- (2) $F^C = F_{1:N} \in \mathbb{R}^{N \times D_F}$: Fitness transformations of children extracted from joint transforms (z-scores & centered ranks).
- (3) $F^P = F_{N:(N+E)} \in \mathbb{R}^{E \times D_F}$: Fitness transformations of parents extracted from joint transforms (z-scores & centered ranks).
- (4) $F^{P'} \in \mathbb{R}^{N \times D_F}$: (Separate) fitness transforms of sampled parents after selection operation (z-scores & centered ranks).
- (5) $\hat{\sigma}^P \in \mathbb{R}^{N \times D_\sigma}$: Mutation strength transformations of sampled parents (z-scores & $[-1, 1]$ normalization).
- (6) $F^M = [\tilde{F}^P, \hat{\sigma}^P] \in \mathbb{R}^{N \times (D_F + D_\sigma)}$: Concatenated fitness and mutation rate features of sampled parents. The fitness features additionally include a Boolean indicating whether an individual performs better than the best fitness observed so far.

Selection via Cross-Attention. We replace the common sorting-based selection mechanism with a cross-SDPA layer, which compares children and parents. It first embeds the fitness transformations of the parents and children into queries, keys and values:

$$Q^P = F^P W_{Q^P} \in \mathbb{R}^{E \times D_K},$$

$$K^C = F^C W_{K^C}, V^C = F^C W_{V^C} \in \mathbb{R}^{N \times D_K}.$$

Afterwards, we compute the normalized dot-product cross-attention features A^S and construct a selection matrix $M^S \in \mathbb{R}^{E \times (N+1)}$:

$$\begin{aligned} A^S &= \text{softmax} \left(\frac{Q^P (K^C)^T}{\sqrt{D_K}} \right) V^C \in \mathbb{R}^{E \times D_K} \\ Q^S &= A^S W_{Q^S} \in \mathbb{R}^{E \times D_K}, K^S = F_C W_{K^S} \in \mathbb{R}^{N \times D_K} \\ M^S &= \text{softmax} \left(\left[\frac{Q^S (K^S)^T}{\sqrt{D_K}}, \mathbf{1}_E \right] \right) \in \mathbb{R}^{E \times (N+1)} \end{aligned}$$

In the final line we concatenate an E -dimensional vector of ones to the outer product of Q^S and K^S . Intuitively, this column represents a fixed offset used to indicate whether the parent copies any child at all or if it is not replaced. The rows of M^S then specify the probability of each offspring to replace a parent:

$$M^S = \text{softmax} \left(\begin{bmatrix} m_{11} & m_{12} & \dots & m_{1N} & 1 \\ m_{21} & m_{22} & \dots & m_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ m_{E1} & m_{E2} & \dots & m_{EN} & 1 \end{bmatrix} \right)$$

M^S is row stochastic, i.e. the rows sum to 1. E.g. M_{11}^S denotes the probability of replacing parent 1 with child 1, while M_{1N+1}^S corresponds to not replacing the first parent. We sample row-wise from a categorical distribution in order to determine whether a child replaces a particular parent. Afterwards, we use the selection matrix to update the parent archive ($X^{P'}$) and fitness archive ($f^{P'}$):

$$\begin{aligned} S &\sim \text{Categorical}(M^S) \text{ with } S \in \mathbb{R}^{E \times (N+1)}, \\ X^{P'} &= X^C \cdot S_{:,1:N} + X^P \cdot \text{diag}(S_{:,N+1}), \end{aligned}$$

and similarly we obtain $f^{P'}$ and $\sigma^{P'}$ via masked addition. This selection operator can flexibly regulate the amount of truncation selection by replacing multiple parent slots with the same child.

Mutation Rate Adaptation (MRA) via Self-Attention. Next to selection we meta-learn MRA. The concatenated fitness and MR features of the sampled parents are processed by a SDPA layer, which outputs a child-specific feature matrix $A^M \in \mathbb{R}^{N \times D_K}$:

$$K^M = F^M W_{K^M}, Q^M = F^M W_{Q^M}, V^M = F^M W_{V^M} \in \mathbb{R}^{N \times D_K},$$

$$A^M = \text{softmax} \left(\frac{Q^M (K^M)^T}{\sqrt{D_K}} \right) V^M \in \mathbb{R}^{N \times D_K}.$$

Afterwards, the multiplicative adaptation to the MR is constructed by projecting and re-parametrizing the attention output:

$$\Delta_\sigma = \exp(0.5 \times A^M W_\sigma) \in \mathbb{R}^N$$

The children MR σ^C is obtained via element-wise multiplication:

$$\sigma^C = \Delta_\sigma \odot \sigma^{P'} \in \mathbb{R}^N$$

$\theta = \{W_{Q^P}, W_{K^C}, W_{V^C}, W_{Q^S}, W_{K^S}, W_{Q^M}, W_{K^M}, W_{V^M}, W_\sigma\}$ denotes the joint attention weights, which characterize a specific instance of an LGA. We use a small feature dimension $D_K = 16$, which results in <1500 trainable meta-parameters. In summary, we introduced two dot-product attention-based operators, which replace the standard selection and MRA operations. Throughout the paper we focus on learned selection and MRA, but in Appendix A we outline how to additionally construct sampling and cross-over operators using self-attention.

5 META-TRAINING, OBJECTIVE & TASKS

We meta-optimize the weights of the GA attention modules to perform BBO on a family of representative tasks. More specifically, we make use of the previously introduced MetaBBO procedure [25] and evolve the LGA parameters to maximize performance on a task distribution of 10 BBOB [13] functions. These include functions with different properties, i.e. separability, conditioning, multi-modality (see Table 1). At each meta-generation (see Figure 1) we start by uniformly sampling a set of BBO tasks and LGA parameters from a meta-evolutionary optimization algorithm. We denote the set of parameters characterizing the LGA by θ_i for $i = 1, \dots, M$ meta-population members. Afterwards, each LGA is evaluated on all tasks by running an inner loop search. We compute an aggregated meta-performance score to update the meta-EO. The *MetaBBO*-

Algorithm 1 MetaBBO Training of Learned Genetic Algorithms

```

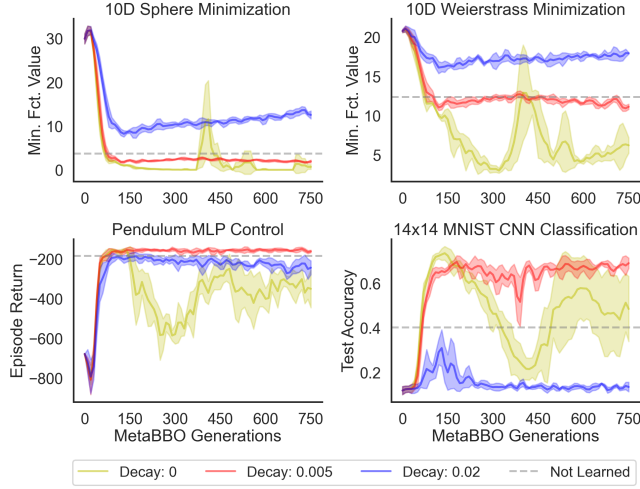
1: Inputs: Meta-population  $M$ , meta-task size  $J$ , inner loop popu-
   lation size  $N$ , generations  $T$ , MetaEO (e.g. OpenAI-ES, [37]).
2: Initialize the meta-search distribution  $\mu, \Sigma = \sigma_{meta} I$ .
3: while not done do
4:   Sample  $J$  BBO tasks with  $\xi_l, \forall l = 1, \dots, J$ .
5:   Sample LGA candidates:  $\theta_i \sim \mathcal{N}(\mu, \Sigma), \forall i = 1, \dots, M$ .
6:   Evaluate all  $M$  LGA candidates on the same  $K$  tasks:
7:   for  $l = 1, \dots, J$  do
8:     for  $i = 1, \dots, M$  do
9:       Initialize search and fitness archives  $X^P, \sigma^P, f^P$ .
10:      for  $t = 0, \dots, T - 1$  do
11:        Sample:  $\tilde{X}^P, \tilde{f}^P, \tilde{\sigma}^P \leftarrow \text{Sample}(X^P, f^P, \sigma^P)$ .
12:        Perform MRA:  $\sigma^C \leftarrow \text{Adaptation}_{\theta_i}(\tilde{f}^P, \tilde{\sigma}^P)$ .
13:        Mutate:  $X^C \leftarrow \text{Mutation}(\tilde{X}^P, \sigma^C)$ .
14:        Evaluate all children:  $f(X_j^C | \xi_l), \forall j = 1, \dots, N$ .
15:        Update parent archive:
16:         $X^{P'}, f^{P'}, \sigma^{P'} \leftarrow \text{Selection}_{\theta_i}(f^C, f^P)$ .
17:      end for
18:    end for
19:  end for
20:  Collect fitness scores  $\left[ \left[ \left[ f(x_{j,t}^C | \xi_l) \right]_{j=1}^N \right]_{l=1}^J \right]_{i=1}^M$ .
21:  Compute normalized & aggregated meta-fitness  $\{\tilde{f}(\theta_i)\}_{i=1}^M$ .
22:  Update meta-search  $\mu', \Sigma' \leftarrow \text{MetaEO}(\{\theta_i, \tilde{f}\}_{i=1}^M | \mu, \Sigma)$ .
23: end while
```

objective is computed based on the collected inner loop fitness scores of each LGA instance, where ξ_l denotes task-specific parameters for $l = 1, \dots, J$ tasks. For each candidate θ_i we minimize the final performance of the best population member. Afterwards, we z-score the task-specific results over meta-population members and compute the median across tasks:

$$\begin{aligned} \left[[f(\theta_i | \xi_k)]_{i=1}^M \right]_{k=1}^K &= \left[\left[\min_j \{f(x_{j,T} | \xi_l)\}_{j=1}^N | \theta_i \right]_{l=1}^J \right]_{i=1}^M \\ \{\tilde{f}(\theta_i)\}_{i=1}^M &= \text{Median} \left[\text{Z-Score} \left([f(\theta_i | \xi_k)]_{i=1}^M \right) \right]_{k=1}^K. \end{aligned}$$

The outer loop optimizes θ using OpenAI-ES [37] for 750 meta-generations with a meta-population size of $M = 512$. We sample

Figure 3: LGA MetaBBO. Meta-evaluation across meta-generations. Evaluation of LGA on two 10 dim. BBOB (Top) and neuroevolution tasks (Bottom). We report mean & 1.96 standard error intervals across 3 independent MetaBBO runs.



$J = 256$ tasks and evaluate each sampled LGA on each task independently. Each LGA is unrolled for $T = 50$ generations, with $N = 16$ population members in each iterations. Throughout, we assume that $E = N$, i.e. the number of parents is equal to the number of children. This is not a limitation since the selection operator is capable of replacing multiple parents with the same child (see Section 7.1). The large-scale parallel meta-evaluation of all θ_i on all ξ_j tasks is facilitated by the auto-vectorization and device parallelism capabilities provided by the JAX library [4, 23] and runs on multiple accelerators. We provide more details and experiments on the meta-training settings and robustness in Appendix B.1 and C.

6 EXPERIMENTS

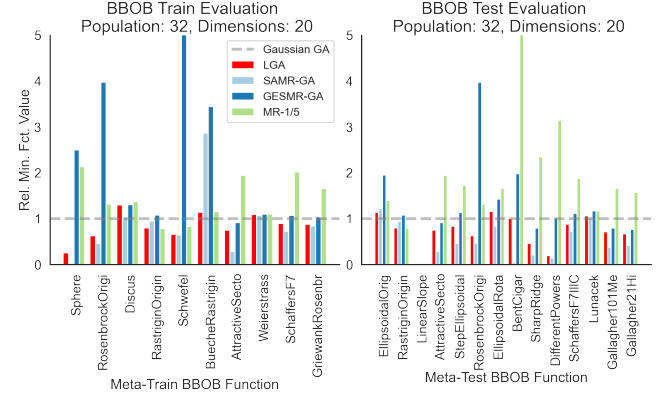
We now turn to an exhaustive experimental evaluation of the MetaBBO optimization procedure and the discovered LGA. We thereby set out to answer the following questions:

- (1) Is it possible to meta-evolve competitive LGAs via MetaBBO using a limited set of meta-training BBO tasks (Section 6.1)?
- (2) Does the resulting LGA outperform GA baselines on unseen BBO problems and different search budgets (Section 6.2)?
- (3) How much can an LGA discovered on a limited set of tasks generalize beyond its meta-training setting (e.g. hyperparameter optimization & neuroevolution; Sections 6.3 & 6.4)?

6.1 Meta-Training on BBOB Functions

We start by meta-evolving the LGA parameters on a task distribution consisting of 10 BBOB functions with different random optima offsets, evaluation noise and considered problem dimensionality ($D \leq 10$). Throughout meta-training we evaluate the performance of the optimized LGA on several different downstream tasks. These include BBOB functions seen during meta-training, hold-out meta-test BBOB functions and small neuroevolution tasks. In Figure 3 we plot the detailed evaluation curves across meta-training. The MetaBBO-trained LGA quickly learns how to perform optimization

Figure 4: Meta-Evaluation on BBOB Tasks. Left: Training Functions. Right: Hold-Out Functions. Scores are normalized by the Gaussian GA baseline performance. Lower is better. Results are averaged over 50 independent evaluation runs.



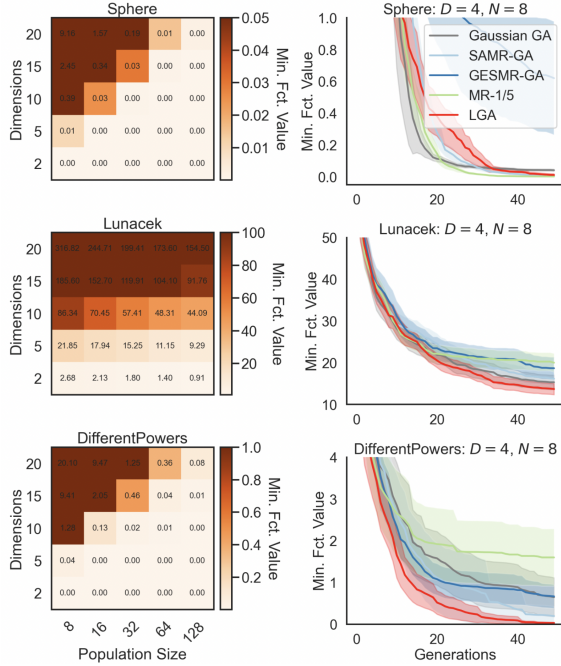
on the low-dimensional BBOB meta-training functions. Interestingly, we find that the MetaBBO procedure can lead to an LGA that overfits to the BBOB tasks on which it was meta-trained. The downstream performance of LGA decreases and becomes unstable for both an unseen Pendulum control task with MLP policy and a down-sized 14-by-14 MNIST classification task using a CNN. We therefore investigated whether meta-regularization can improve the generalization on such unseen neuroevolution tasks. More specifically, we compared three different meta-mean regularization coefficients $\lambda \in \{0, 0.005, 0.02\}$, which exponentially decay the meta-mean to zero, $\mu'_\lambda = (1 - \lambda)\mu'$. We observe that the generalization to the neuroevolution tasks can be improved and stabilized using a properly chosen decay of 0.005. In Appendix C we further explore the impact of the meta-task distribution, meta-objective and LGA attention size. The MetaBBO procedure is largely robust to the choice of these settings. Small attention layers are sufficient for consistently discovering performant LGAs. This comes with the additional advantage of reducing the FLOPs and memory requirements of executing the LGA. The evaluation of a meta-trained LGA is easily feasible on a single core CPU device.

6.2 Meta-Testing on BBOB Functions

Next, we exhaustively evaluate the performance of LGA on the full set of BBOB benchmark functions including test functions unseen during meta-training. We compare against 4 competitive GA baselines: Gaussian GA, MR-1/5 GA, SAMR-GA, GESMR-GA.¹ We compare the performance on all BBOB functions for a population size of $N = 32$, $D = 20$ search dimensions and for $T = 50$ generations. The best-across generations function value is normalized by the performance of the Gaussian GA. In Figure 4 we find that LGA outperforms all baselines on the majority of both BBOB functions seen during meta-training (left) and unseen BBOB functions (right). This holds true for functions with very different characteristics (single/multi-modal, high/low conditioning, separable/not separable), search dimensions and population sizes (Appendix Figure 17 & 18). This provides further evidence that LGA does in fact not

¹All baselines were tuned using grid searches over the parent archive size and initial mutation rate scale. We report all BBOB evaluation & tuning settings in Appendix B.2.

Figure 5: LGA Generalization. Left: BBOB evaluation for different budgets & search spaces. Right: Performance across generations. We report mean & 1.96 standard error intervals across 50 runs.

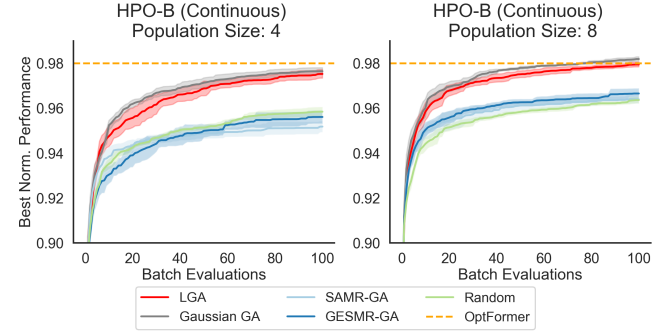


overfit to the BBO functions seen during meta-training, but instead has discovered a general-purpose GA algorithm. In Figure 5 we further demonstrate that LGA generalizes to different population sizes and problem dimensions. The meta-learned GA achieves lower function values in fewer generations (right) and performs well for different problem settings (left). As the problems become harder with increased dimensionality, LGA can compensate with a larger population size.

6.3 Meta-Testing on Continuous HPO-B

Next, we test LGA's performance on the HPO-B benchmark [2]. The benchmark considers a vast array of hyperparameter optimization tasks including 16 different model types (SVM, XGBoost, etc.) and their respective search spaces ($D \in \{2, 3, \dots, 16\}$). Each model is evaluated on 2 to 7 different datasets, which leads to a total of 86 hyperparameter search tasks. We consider the continuous HPO-B version, which uses a previously fitted surrogate model. Note that the LGA has not been trained on such hyperparameter optimization tasks. Figure 6 compares the performance of LGA against the GA and a random search baselines. Additionally, we report the reference performance of the recently proposed OptFormer model [6] after 105 total evaluations. We find that LGA outperforms the majority of considered GA baselines. Again, this observation holds for two considered population sizes. LGA can also achieve similar performance as OptFormer, which has been trained on a much more diverse task distribution. This highlights the transfer capabilities and applicability of LGA to new during meta-training unseen optimization domains.

Figure 6: LGA Evaluation on HPO-B [2]. Left: Small population size ($N = 4$). Right: Large population size ($N = 8$). We report mean & 1.96 standard error intervals across 5 runs.



6.4 Meta-Testing on Neuroevolution Tasks

Until now we have evaluated the performance of the discovered LGA on moderately small search spaces (i.e. $D \leq 20$). But is it also possible to deploy LGA to neuroevolution settings with thousands of search dimensions and arguably very different fitness landscape characteristics? Again, note that LGA has never explicitly been trained to evolve such high-dimensional genomes and that this requires strong transfer of the learned GA operators.

The considered Reinforcement Learning (RL) tasks consist of 4 robotic control tasks (Pendulum-v1 [24] and 5 Brax tasks [10]) using MLP policies and three MinAtar visual control tasks (SpaceInvaders, Breakout & Asterix [46]) with CNN-based policies. The MLP genomes consist of less than 1000 weights, while the MinAtar CNNs have ca. 50,000. Hence, the search space is many orders higher than what the LGA has been meta-trained on ($D \leq 10$). The top three rows of Figure 7 show that LGA can compete with all tuned baseline GAs on the nine RL tasks with different search spaces, fitness landscapes and evaluation budgets. Interestingly, the performance gap between LGA and the considered baselines is the biggest for the CNN policies and tends to increase with the number of search dimensions. Finally, in the final row of Figure 7 we show that LGA can also successfully be applied to three image classification tasks including MNIST, Fashion-MNIST and K-MNIST classification (28-by-28 grayscale images) with a small CNN (2 convolutional layers, ReLU activation and a linear readout) with 11274 evolvable weights. The LGA generalizes far beyond the meta-training search horizon ($T = 50$ versus 4000) and does not meta-overfit [26].

7 ANALYSIS OF DISCOVERED LGA

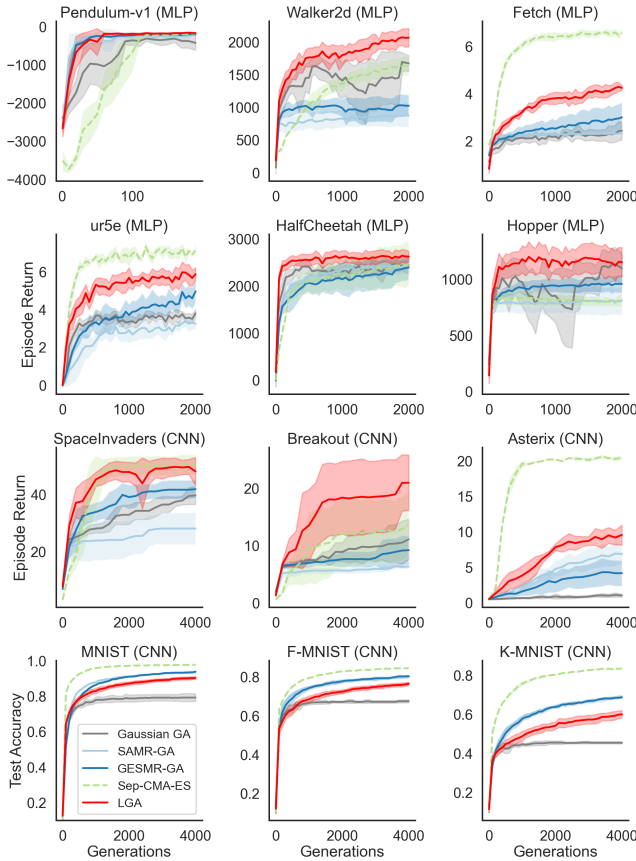
After having established that the meta-trained LGA is capable of outperforming a set of GA baselines on unseen optimization problems, we now investigate the underlying discovered mechanisms, transfer ability and robustness of the meta-learned GA operators.

7.1 Visualization of Learned Genetic Operators

What types of mechanisms underlying the black-box genetic operators has LGA discovered? Has it simply re-discovered fitness-based truncation selection or a more complex parent replacement procedure? In Figure 8 we consider a 2-dim Sphere problem and visualize the selection mask $S_{:,1:N}$ used to update the parent archive X^P . We observe that the selection operator uses children solution to replace parents based on their improvement over the best seen solution.

Furthermore, one well-performing child often times replaces more than a single parent. This indirectly implies that the selection operator has meta-learned to dynamically adapt its elite archive size and thereby also the effective sampling distribution. A child that has replaced multiple parents will be sampled (with replacement) more frequently in the next generation. Furthermore, this implies a robustness mechanism: Since children can be stored multiple times in the parent archive, they are less likely to be ‘forgotten’ by the stochastic selection. The mutation rates, on the other hand, are decreased over the course of generations in order to explore closer to the global optimum. Furthermore, we observe a grouping of the MR based on the performance of the parents. Children with bad performing parents tend to exhibit a higher mutation rate.

Figure 7: LGA Evaluation on neuroevolution tasks including continuous control (top), visual control (middle) & computer vision (bottom) tasks. We report mean & 1.96 standard error intervals across 5 independent runs.

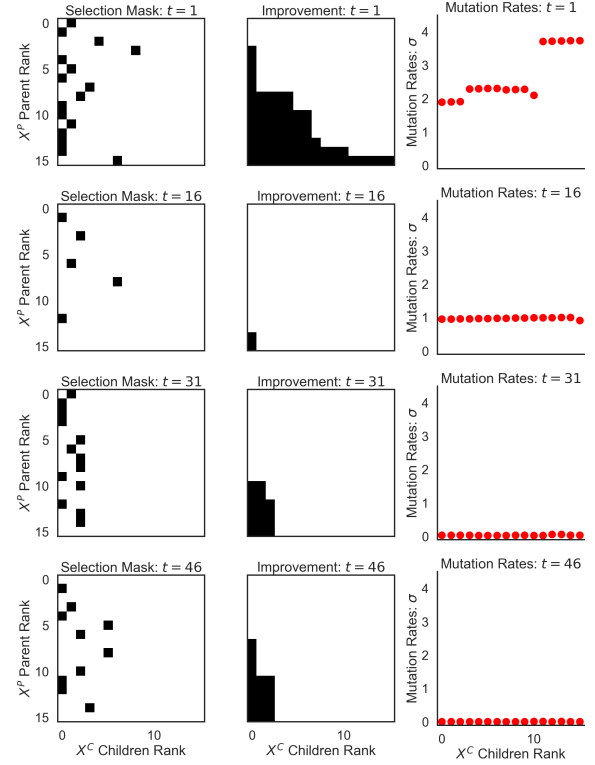


7.2 Ablation & Transfer of Genetic Operators

How much do the different learned components contribute to the overall performance of LGA? Can the learned modules act as drop-in replacements for other genetic algorithms? To answer this question we consider two types of comparative studies:

- (1) **Operator ablation before MetaBBO discovery:** We meta-train the LGA with a variable amount of learned genetic operators. E.g. we fix the selection operator to white-box

Figure 8: LGA’s selection operator on a 2-dim Sphere task. Left: Sampled selection matrix $S_{:,1:N} \in \mathbb{R}^{E \times N}$. Middle: Matrix indicating whether a child improves the fitness score over the parents. Right: Mutation rates for different children. Rows indicate 4 different generations $t \in \{1, 16, 31, 46\}$.



truncation selection and only meta-learn mutation rate adaptation. This allows us to quantify the joint contributions and synergies between the different learned ingredients.

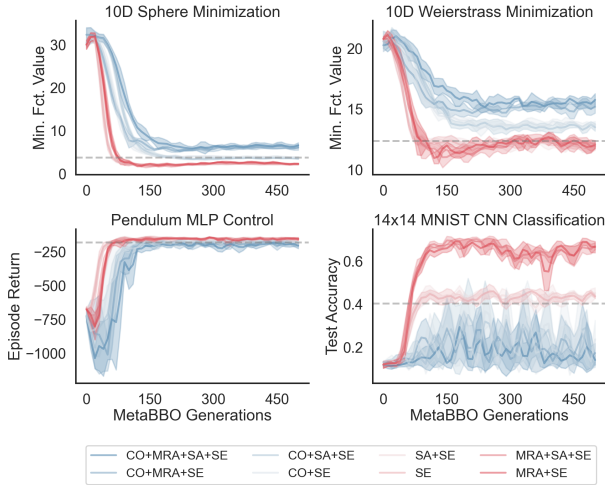
- (2) **Operator transfer after MetaBBO discovery:** After meta-training is completed, we ask whether or not it is possible to substitute the learned operators into other genetic algorithms? This in turn allows us to assess whether the specific learned operator is overfit to the downstream GA computations or whether it can act as a transferable inductive bias for genetic computation.

For the first study we compare meta-training combinations of attention-parametrized selection (SE), mutation rate adaptation (MRA), cross-over (CO) and sampling (SA).² For each combination we plot the evaluation performance across meta-generations in Figure 9. We observe that MRA is crucial for good performance on the neuroevolution tasks. Intuitively, this can be explained by the smaller scale of solution parameters associated with neural network weights. The GA benefits from the ability to flexibly down-regulate its perturbation strengths. Cross-over, on the other hand, is detrimental for the generalization of LGA to the MNIST CNN

²CO and SA parametrizations are additionally introduced in Appendix A. Throughout the main text we mainly focused on learned mutation rate adaptation and selection.

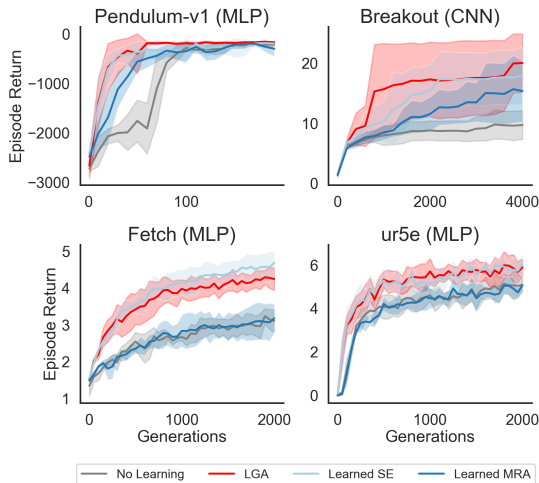
neuroevolution task. This behavior can arguably be attributed to the challenge of finding beneficial crossing over pairs for different neural network genomes. We further observed that learned sampling does not significantly improve the performance of the LGA. We hypothesize that this is due to the indirect effect of the selection mechanism on the sampling of children. Finally, the overall best performing configuration only meta-learns selection and MRA.

Figure 9: Visualization of LGA’s operator ablations during MetaBBO. Evaluation of LGA on two 10 dim. BBOB (Top) and neuroevolution tasks (Bottom). We report mean & 1.96 standard error intervals across 3 independent MetaBBO runs.



Next, we considered replacing the truncation selection and fixed mutation rate of the Gaussian GA baseline with the learned selection and MRA operators. In Figure 10 we show that this can successfully be accomplished for four neuroevolution tasks. Replacing either the selection or adding learned MRA operator improves the performance of the Gaussian GA. The learned operators can act as drop-in replacements and are transferable inductive biases.

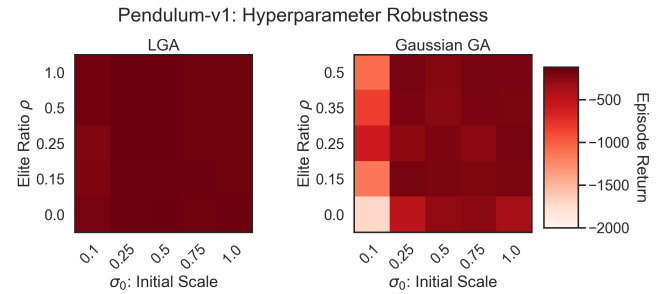
Figure 10: Transfer of learned operators to a Gaussian GA. We report mean & 1.96 standard error intervals across 5 runs.



7.3 Hyperparameter Robustness of LGA

Finally, we assess the sensitivity of LGA to its remaining hyperparameter choices. More specifically, we compare the performance of LGA and the baseline GAs for various initial mutation rate scales σ_0 and parent archive sizes $E = \lceil \rho \times N \rceil$, where ρ denotes the fraction of population members making up the number of parents. Note that while LGA was meta-trained for $\rho = 1$, i.e. $E = N$, we find that it is capable of generalizing to many different archive sizes and is robust to the initial scale parameters (Pendulum control task; see Figure 11). In Section D.2 we provide the same analysis for all neuroevolution tasks. LGA is far less hyperparameter sensitive than the considered baseline GAs. This highlights the robustness of the LGA induced by the MetaBBO process.

Figure 11: Hyperparameter Robustness of LGA. Pendulum-v1 performance across elite ratios and initial mutation rates. $\rho = 0$ uses a single parent $E = 1$. Results are averaged over 5 independent evaluation runs.



8 CONCLUSION

Summary. In this study, we used evolutionary optimization to discover novel genetic algorithms via meta-learning. We leveraged the insight that GA operators perform set operations and parametrized them with novel attention modules that induce the inductive bias of permutation equivariance. Our benchmark results on BBOB, HPO-B and neuroevolution tasks highlight the potential of combining flexible GA parametrization with data-driven meta-evolution.

Limitations. We use powerful neural network layers to characterize GA operators. While flexible and interpretable (Section 7.1), the underlying mechanisms do remain partially opaque. Future work needs to be done in order to fully reverse-engineer the discovered operators. In Appendix E we provide a first set of insights unraveling simple linear relationships between the attention inputs and their outputs. We believe these can guide the design of new ‘white-box’ GAs informed by ‘grey-box’ discovered LGAs. Furthermore, our analysis highlights the importance of meta-regularization and the potential for the automated design of meta-training curricula.

Future Work. We are interested in explicitly regularizing LGAs to maintain diversity in their parent archive. This may provide a bridge to meta-learned quality-diversity methods [8]. Furthermore, it may be possible to parametrize a flexible EO algorithm that can interpolate between the exploration of multiple solution candidates as in GA and a single search distribution mode as in traditional evolution strategies. Finally, we believe that better meta-learned GAs can be discovered by simultaneously co-evolving the meta-task distribution and the learned GA.

REFERENCES

- [1] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. 2016. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems* 29 (2016).
- [2] Sebastian Pineda Arango, Hadi S Jomaa, Martin Wistuba, and Josif Grabocka. 2021. Hpo-b: A large-scale reproducible benchmark for black-box hpo based on openml. *arXiv preprint arXiv:2106.06257* (2021).
- [3] Samy Bengio, Yoshua Bengio, Jocelyn Cloutier, and Jan Gescsei. 1992. On the optimization of a synaptic learning rule. In *Optimality in Biological and Artificial Networks?* Routledge, 281–303.
- [4] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. 2018. JAX: composable transformations of Python+NumPy programs. (2018). <http://github.com/google/jax>
- [5] Yutian Chen, Matthew W. Hoffman, Sergio Gómez Colmenarejo, Misha Denil, Timothy P. Lillicrap, Matt Botvinick, and Nando de Freitas. 2017. Learning to Learn without Gradient Descent by Gradient Descent. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, 748–756. <https://proceedings.mlr.press/v70/chen17e.html>
- [6] Yutian Chen, Xingyou Song, Chansoo Lee, Zi Wang, Qiuyi Zhang, David Dohan, Kazuya Kawakami, Greg Kochanski, Arnaud Doucet, Marc'aurelio Ranzato, et al. 2022. Towards Learning Universal Hyperparameter Optimizers with Transformers. *arXiv preprint arXiv:2205.13320* (2022).
- [7] Jeff Clune, Dusan Misevic, Charles Ofria, Richard E Lenski, Santiago F Elena, and Rafael Sanjuán. 2008. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. *PLoS Computational Biology* 4, 9 (2008), e1000187.
- [8] Antoine Cully and Yiannis Demiris. 2017. Quality and diversity optimization: A unifying modular framework. *IEEE Transactions on Evolutionary Computation* 22, 2 (2017), 245–259.
- [9] Sebastian Flennerhag, Yannick Schroeder, Tom Zahavy, Hado van Hasselt, David Silver, and Satinder Singh. 2021. Bootstrapped meta-learning. *arXiv preprint arXiv:2109.04504* (2021).
- [10] C Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. 2021. Brax—A Differentiable Physics Engine for Large Scale Rigid Body Simulation. *arXiv preprint arXiv:2106.13281* (2021).
- [11] C Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. 2021. Brax—A Differentiable Physics Engine for Large Scale Rigid Body Simulation. *arXiv preprint arXiv:2106.13281* (2021).
- [12] Hugo Siqueira Gomes, Benjamin Léger, and Christian Gagné. 2021. Meta learning black-box population-based optimizers. *arXiv preprint arXiv:2103.03526* (2021).
- [13] Nikolaus Hansen, Anne Auger, Steffen Finck, and Raymond Ros. 2010. *Real-parameter black-box optimization benchmarking 2010: Experimental setup*. Ph.D. Dissertation. INRIA.
- [14] Nikolaus Hansen, Steffen Finck, Raymond Ros, and Anne Auger. 2009. *Real-parameter black-box optimization benchmarking 2009: Noisy functions definitions*. Ph.D. Dissertation. INRIA.
- [15] Nikolaus Hansen and Andreas Ostermeier. 2001. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation* 9, 2 (2001), 159–195.
- [16] Charles R Harris, K Jarrod Millman, Stéfán J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. 2020. Array programming with NumPy. *Nature* 585, 7825 (2020), 357–362.
- [17] John D Hunter. 2007. Matplotlib: A 2D graphics environment. *Computing in science & engineering* 9, 03 (2007), 90–95.
- [18] Louis Kirsch, Sebastian Flennerhag, Hado van Hasselt, Abram Friesen, Junhyuk Oh, and Yutian Chen. 2022. Introducing symmetries to black box meta reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 7202–7210.
- [19] Louis Kirsch and Jürgen Schmidhuber. 2021. Meta learning backpropagation and improving it. *Advances in Neural Information Processing Systems* 34 (2021), 14122–14134.
- [20] Jannik Kossen, Neil Band, Clare Lyle, Aidan N Gomez, Thomas Rainforth, and Yarin Gal. 2021. Self-attention between datapoints: Going beyond individual input-output pairs in deep learning. *Advances in Neural Information Processing Systems* 34 (2021), 28742–28756.
- [21] Akarsh Kumar, Bo Liu, Risto Miikkulainen, and Peter Stone. 2022. Effective Mutation Rate Adaptation through Group Elite Selection. *arXiv preprint arXiv:2204.04817* (2022).
- [22] Robert Tjarko Lange. 2021. MLE-Infrastructure: A Set of Lightweight Tools for Distributed Machine Learning Experimentation. (2021). <http://github.com/mle-infrastructure>
- [23] Robert Tjarko Lange. 2022. evosax: JAX-based Evolution Strategies. (2022). <http://github.com/RobertTLange/evosax>
- [24] Robert Tjarko Lange. 2022. gymnax: A JAX-based Reinforcement Learning Environment Library. (2022). <http://github.com/RobertTLange/gymnax>
- [25] Robert Tjarko Lange, Tom Schaul, Yutian Chen, Tom Zahavy, Valenti Dallibard, Chris Lu, Satinder Singh, and Sebastian Flennerhag. 2022. Discovering Evolution Strategies via Meta-Black-Box Optimization. *arXiv preprint arXiv:2211.11260* (2022).
- [26] Robert Tjarko Lange and Henning Sprekeler. 2022. Learning not to learn: Nature versus nurture in silico. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 7290–7299.
- [27] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. 2019. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*. PMLR, 3744–3753.
- [28] Chris Lu, Jakub Grudzien Kuba, Alistair Letcher, Luke Metz, Christian Schroeder de Witt, and Jakob Nicolaus Foerster. 2022. Discovered Policy Optimisation. In *Decision Awareness in Reinforcement Learning Workshop at ICML 2022*.
- [29] Luke Metz, James Harrison, C Daniel Freeman, Amil Merchant, Lucas Beyer, James Bradbury, Naman Agrawal, Ben Poole, Igor Mordatch, Adam Roberts, et al. 2022. VeLO: Training Versatile Learned Optimizers by Scaling Up. *arXiv preprint arXiv:2211.09760* (2022).
- [30] Luke Metz, Niru Maheswaranathan, C Daniel Freeman, Ben Poole, and Jascha Sohl-Dickstein. 2020. Tasks, stability, architecture, and compute: Training more effective learned optimizers, and using them to train themselves. *arXiv preprint arXiv:2009.11243* (2020).
- [31] Luke Metz, Niru Maheswaranathan, Jeremy Nixon, Daniel Freeman, and Jascha Sohl-Dickstein. 2019. Understanding and correcting pathologies in the training of learned optimizers. In *International Conference on Machine Learning*. PMLR, 4556–4565.
- [32] Andrew Y Ng and Michael I Jordan. 2000. PEGASUS: A policy search method for large MDPs and POMDPs. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)*. 406–415.
- [33] Junhyuk Oh, Matteo Hessel, Wojciech M Czarnecki, Zhongwen Xu, Hado P van Hasselt, Satinder Singh, and David Silver. 2020. Discovering reinforcement learning algorithms. *Advances in Neural Information Processing Systems* 33 (2020), 1060–1070.
- [34] Jack Parker-Holder, Raghu Rajan, Xingyou Song, André Biedenkapp, Yingjie Miao, Theresa Eimer, Baohe Zhang, Vu Nguyen, Roberto Calandra, Aleksandra Faust, et al. 2022. Automated reinforcement learning (autorl): A survey and open problems. *Journal of Artificial Intelligence Research* 74 (2022), 517–568.
- [35] Ingo Rechenberg. 1973. Evolutionsstrategie. *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (1973).
- [36] Raymond Ros and Nikolaus Hansen. 2008. A simple modification in CMA-ES achieving linear time and space complexity. In *International conference on parallel problem solving from nature*. Springer, 296–305.
- [37] Tim Salimans, Jonathan Ho, Xi Chen, Shyam Sidor, and Ilya Sutskever. 2017. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864* (2017).
- [38] Gresa Shala, André Biedenkapp, Noor Awad, Steven Adriaensen, Marius Lindauer, and Frank Hutter. 2020. Learning step-size adaptation in CMA-ES. In *International Conference on Parallel Problem Solving from Nature*. Springer, 691–706.
- [39] Yujin Tang and David Ha. 2021. The sensory neuron as a transformer: Permutation-invariant neural networks for reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 22574–22587.
- [40] Yujin Tang, Yingtao Tian, and David Ha. 2022. EvoJAX: Hardware-Accelerated Neuroevolution. *arXiv preprint arXiv:2202.05008* (2022).
- [41] Vishnu TV, Pankaj Malhotra, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. 2019. Meta-learning for black-box optimization. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 366–381.
- [42] Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Remi Munos, Charles Blundell, Dharmash Kumar, and Matt Botvinick. 2016. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763* (2016).
- [43] Michael L Waskom. 2021. Seaborn: statistical data visualization. *Journal of Open Source Software* 6, 60 (2021), 3021.
- [44] Zhongwen Xu, Hado P van Hasselt, Matteo Hessel, Junhyuk Oh, Satinder Singh, and David Silver. 2020. Meta-gradient reinforcement learning with an objective discovered online. *Advances in Neural Information Processing Systems* 33 (2020), 15254–15264.
- [45] Zhongwen Xu, Hado P van Hasselt, and David Silver. 2018. Meta-gradient reinforcement learning. *Advances in neural information processing systems* 31 (2018).
- [46] Kenny Young and Tian Tian. 2019. Minatar: An atari-inspired testbed for thorough and reproducible reinforcement learning experiments. *arXiv preprint arXiv:1903.03176* (2019).
- [47] Tom Zahavy, Zhongwen Xu, Vivek Veeriah, Matteo Hessel, Junhyuk Oh, Hado P van Hasselt, David Silver, and Satinder Singh. 2020. A self-tuning actor-critic algorithm. *Advances in Neural Information Processing Systems* 33 (2020), 20913–20924.