Evolution Strategies for Neural Policy Search

Paul Templier

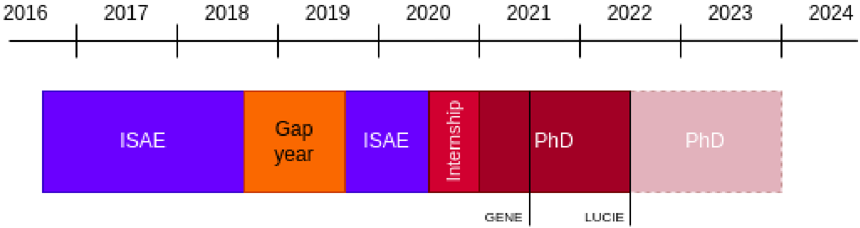
ISAE Supaero, Département Ingéniérie des Systèmes Complexes (DISC)

Institut Supérieur de l'Aéronautique et de l'Espace
SUPAERO

June 29, 2022

Plan

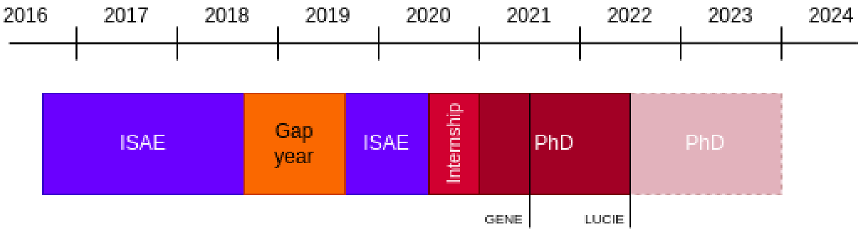
Mid-thesis report



Mid-thesis report

Initial topic

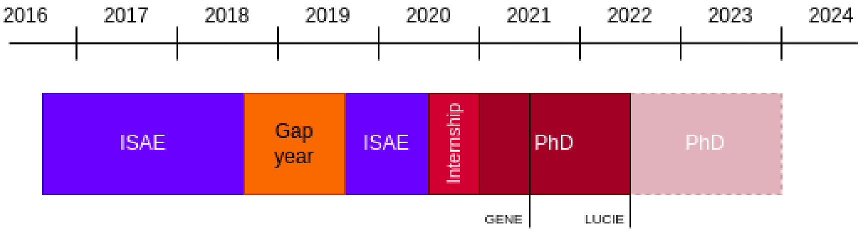
Bio-inspired methods for artificial neural networks



Mid-thesis report

Initial topic

Bio-inspired methods for artificial neural networks



Goal of this report

Organize past and present work, and highlight future research directions.

Content

1

Content

1.
2.

Content

1.
2.
3.

Content

1.
2.
3.
4.

Content

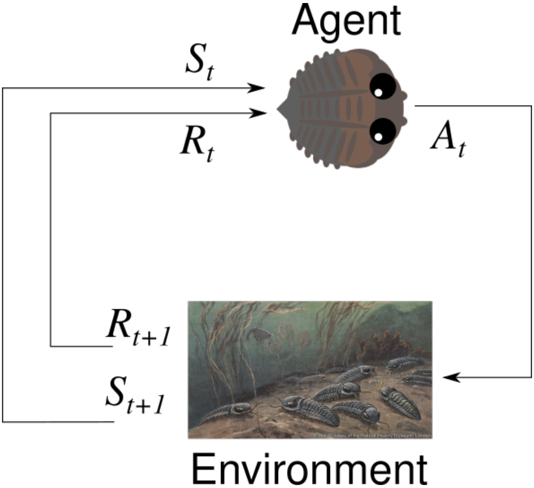
1.
2.
3.
4.
5.

Content

1.
2.
3.
4.
5.
6.

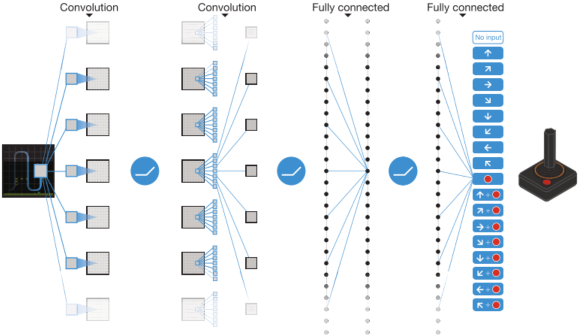
Policy search

Policy search

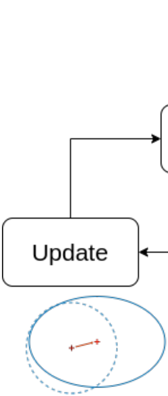


https://github.com/d9w/evolution/blob/master/imgs/erl.png

Neural networks



Neural Network used in Deep Q Networks [?]



Evolution Strategy steps

Variants of

Evolution Strategies

- (μ, λ) ES
- SNES
- Canonical ES
- OpenAI ES
- CMA-ES
- XNES
- Cross-Entropy Method
- Augmented Random Search

Variants of

Evolution Strategies

- (μ, λ) ES
- SNES
- Canonical ES
- OpenAI ES
- CMA-ES
- XNES
- Cross-Entropy Method
- Augmented Random Search

Neuroevolution for policy search

- large dimensions ($1.6\ .10^6$ parameters)
- expensive evaluation

Reproduction settings

Reproducing Canonical ES [?] and OpenAI ES [?] on the Arcade Learning Environment.

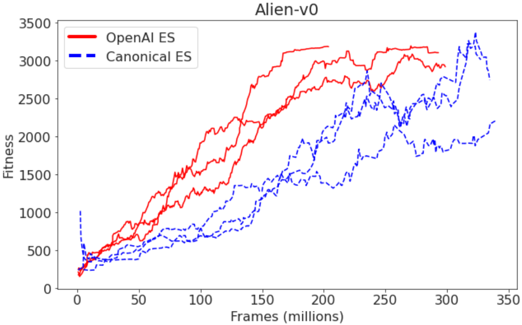
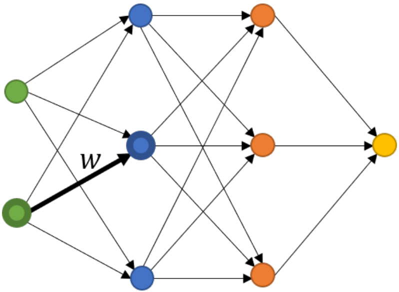


Figure: Evolution of Canonical ES and OpenAI ES on Alien with 800 CPUh compute budget

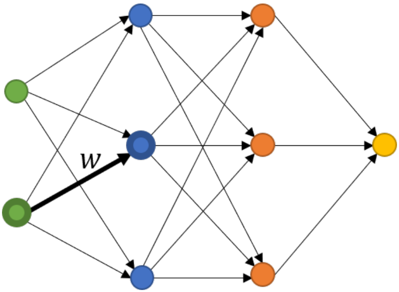
A Geometric Encoding for Neural Network Evolution

A Geometric Encoding for Neural Network Evolution

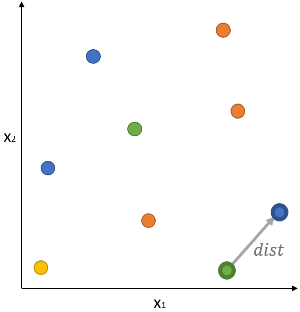


Fully connected
neural network

A Geometric Encoding for Neural Network Evolution



Fully connected
neural network



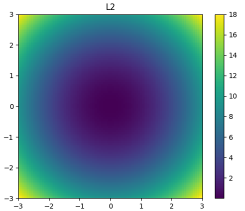
GENE encoding

: Distance functions

$$w_{i,j} = dist(n_i, n_j) \tag{1}$$

Euclidean distance

$$\sqrt{\sum_{k=1}^{D} \left(n_1^k - n_2^k\right)^2} \tag{2}$$



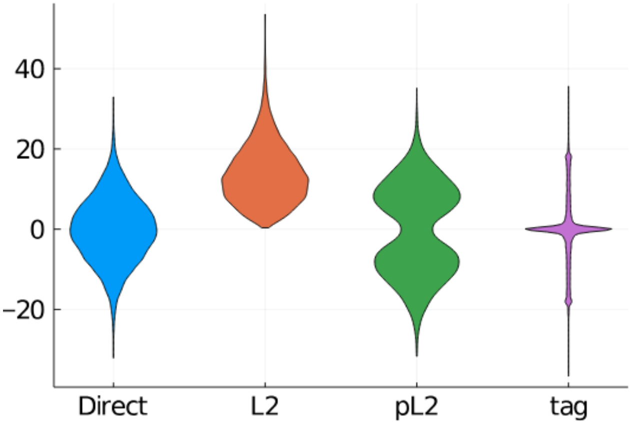


Figure: Distribution of weight values in networks evolved with different encodings.

Competitive results - Arcade Learning Environment

Competitive results - Arcade Learning Environment

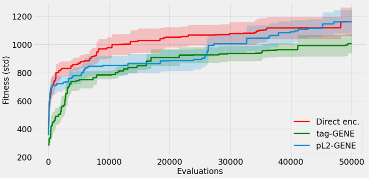


Figure: SNES on SpaceInvaders

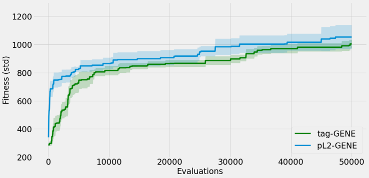


Figure: XNES on SpaceInvaders

Competitive results - Arcade Learning Environment

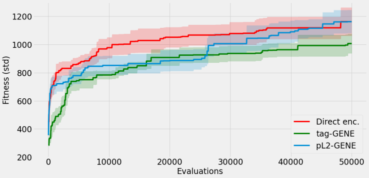


Figure: SNES on SpaceInvaders

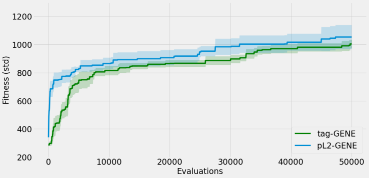


Figure: XNES on SpaceInvaders

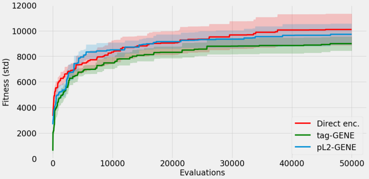


Figure: SNES on Krull

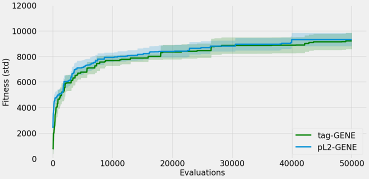


Figure: XNES on Krull

Improving results - Arcade Learning Environment

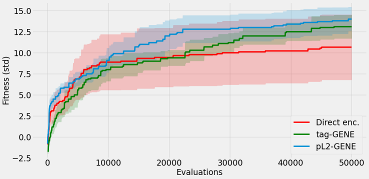


Figure: SNES on IceHockey

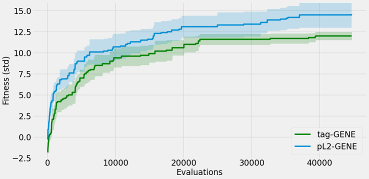


Figure: XNES on IceHockey

Improving results - Arcade Learning Environment

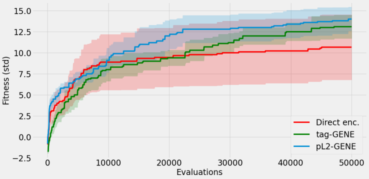


Figure: SNES on IceHockey

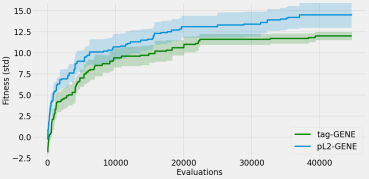


Figure: XNES on IceHockey

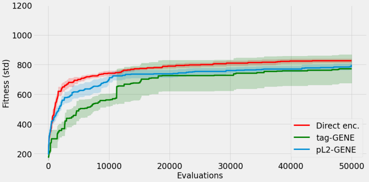


Figure: SNES on Seaquest

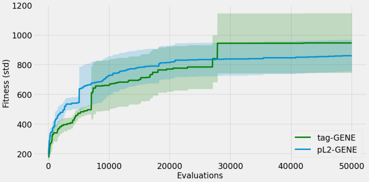


Figure: XNES on Seaquest

Computational cost

Evolutionary Strategy update of μ and σ

Encoding	D	Genes		Mean time (s)	Memory (KiB)
pL2-GENE	3	804	SNES	0.000357	630.56
pL2-GENE	10	2211	SNES	0.000678	1372.16
Direct	-	5609	SNES	0.001350	3133.44
pL2-GENE	3	804	XNES	1.475000	1352663.04
pL2-GENE	10	2211	XNES	14.244000	11806965.76
Direct	-	5609	XNES	119.976000	79765176.32

Distance functions

Design new distance functions, or optimize them through co-evolution.

Gradient descent

Use backpropagation and gradient descent to optimize genomes instead of evolution.

Hybrid encoding

Switch between indirect and direct encodings during the evolution.

Complex networks

Design encodings for convolution layers and recurrent networks.

ES on noisy environments

ES on noisy environments

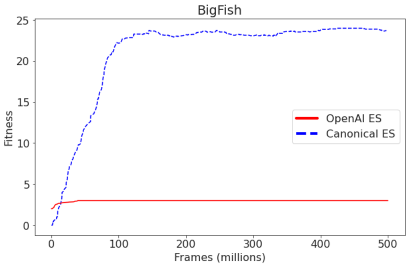


Figure: ES on BigFish, same level

ES on noisy environments

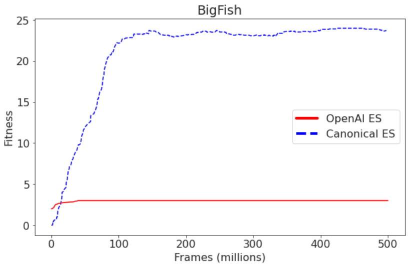


Figure: ES on BigFish, same level

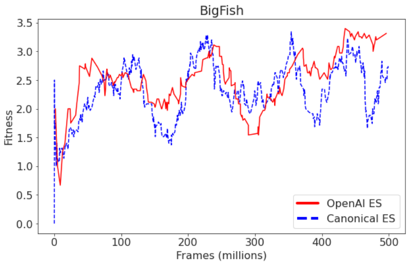


Figure: ES on BigFish, random level

The selection procedure

The selection procedure

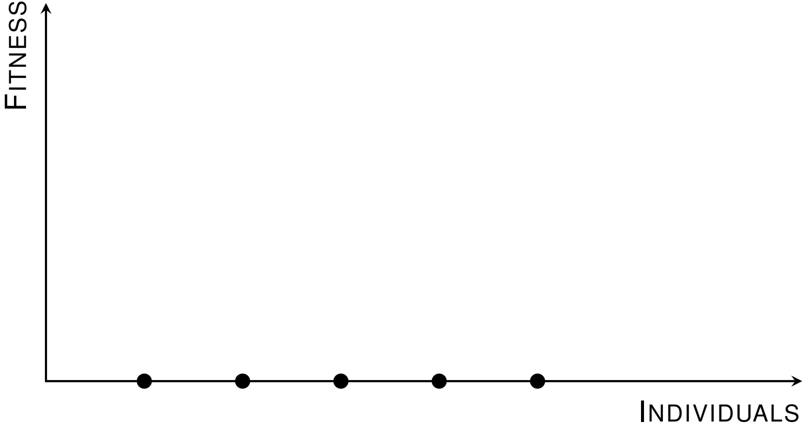
Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]

The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]

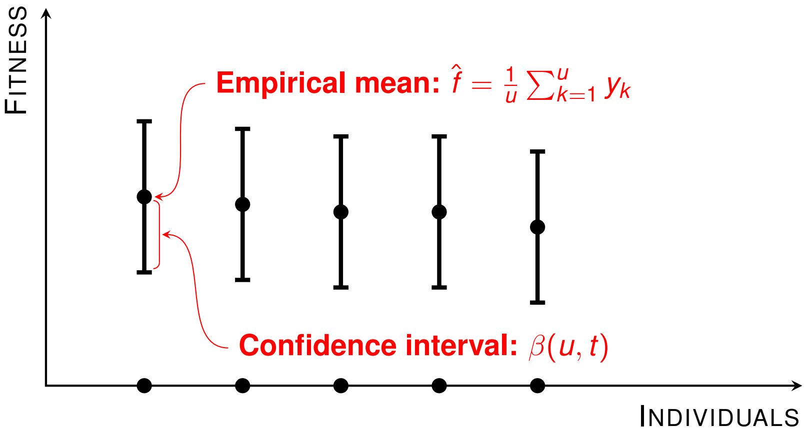
The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]



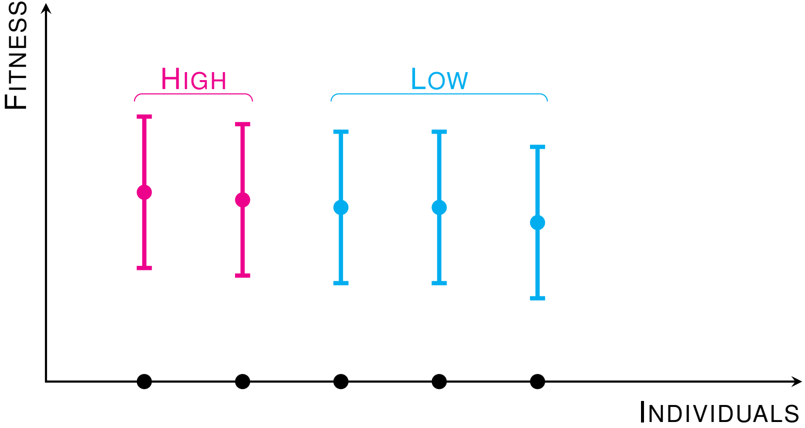
The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]



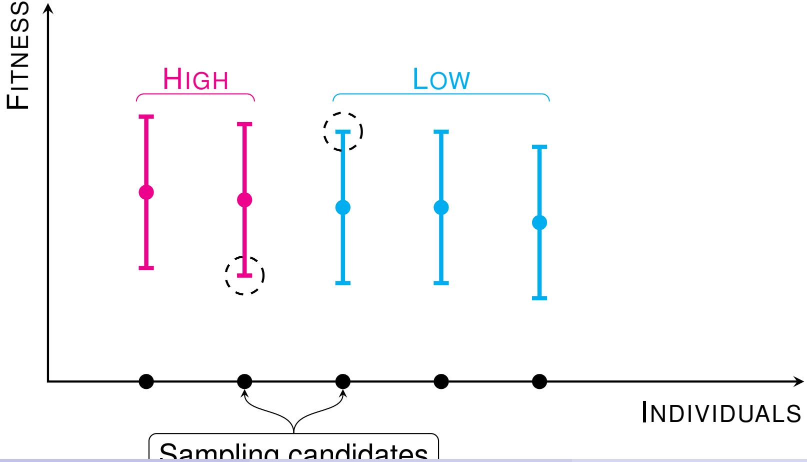
The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]



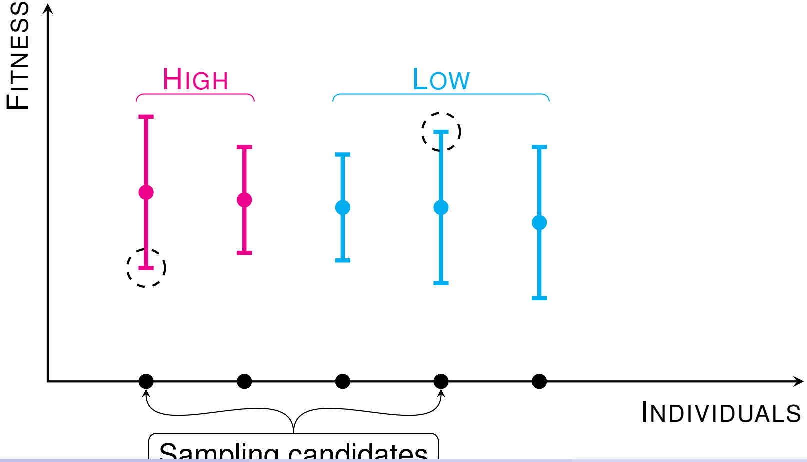
The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]



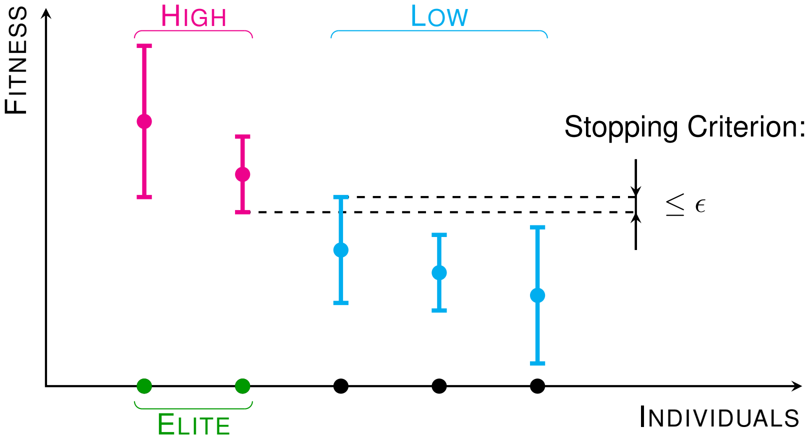
The selection procedure

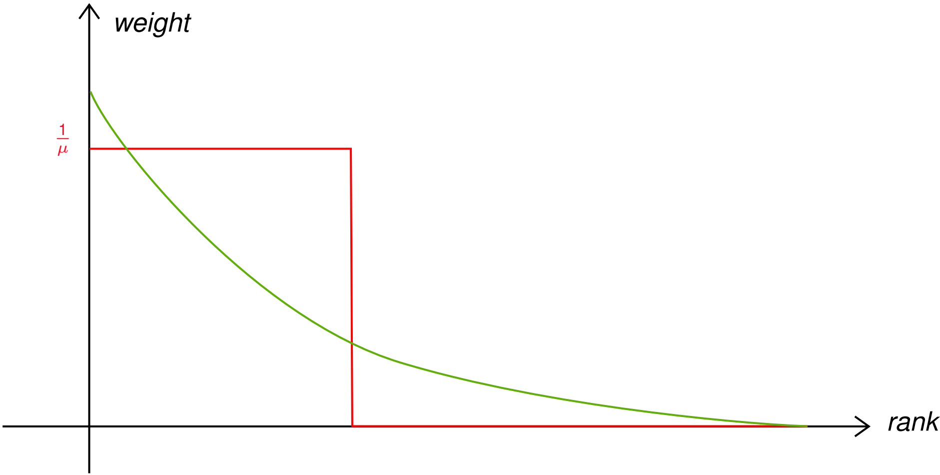
Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]



The selection procedure

Objective: identify the **best** μ individuals with as **few evaluations** as possible. [**?**]

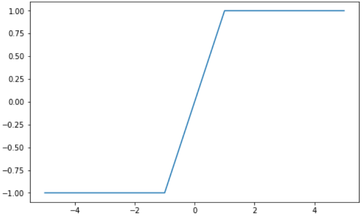



weight
$\frac{1}{\mu}$
rank

Signed distances

Bounded identity function

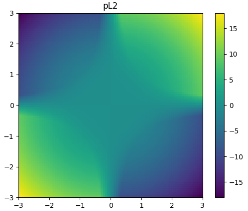
$$\alpha : \begin{cases} \text{if } x \geq 1 : \alpha(x) = 1 \\ \text{if } x \leq -1 : \alpha(x) = -1 \\ \text{else: } \alpha(x) = x \end{cases} \tag{3}$$



Distance functions

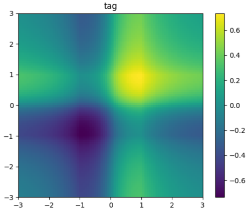
pL2-GENE

$$\alpha \left(\prod_{k=1}^{D} n_1^k - n_2^k \right) \sqrt{\sum_{j=1}^{D} \left(n_1^j - n_2^j \right)^2} \qquad (4)$$



tag-GENE

$$\sum_{j=2}^{D} \alpha (n_1^j - n_2^1) e^{-|n_1^j - n_2^1|} \qquad (5)$$



References I

P. Chrabaszcz, I. Loshchilov, and F. Hutter.
Back to Basics: Benchmarking Canonical Evolution Strategies for Playing Atari.
pages 1419–1426, 2018.

T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever.
Evolution Strategies as a Scalable Alternative to Reinforcement Learning.
Mar. 2017.