

---

RAPPORT DE PROJET :

BACK TO BASICS : BENCHMARKING CANONICAL EVOLUTION  
STRATEGIES FOR PLAYING ATARI

---

MU5IN259 - Intelligence Artificielle pour la Robotique

M2 Informatique - Spécialité ANDROIDE



Année universitaire 2019-2020

## Travail réalisé

L'objet de ce projet était tout d'abord de reproduire les résultats obtenus par Patryk Chrabaszcz, Ilya Loshchilov et Frank Hutter dans leur article *Back to Basics : Benchmarking Canonical Evolution Strategies for Playing Atari* sur un seul jeu Atari, avec un budget plus limité et des réseaux de neurones de plus petite taille. Dans un premier temps nous avons alors comparé les performances des algorithmes OpenAI ES et Canonical ES dans l'apprentissage du jeu Pong. Nous avons ensuite étendu cette comparaison aux algorithmes CEM et CMA-ES et DQN.

## Description du jeu

Le jeu Atari Pong est une simulation simpliste de tennis de table (ping-pong) dans lequel deux joueurs s'affrontent jusqu'à ce que l'un d'eux atteignent 21 points. Chaque joueur commande une raquette représentée par un trait vertical et doit renvoyer la balle dans le camp de son adversaire. S'il n'y parvient pas (i.e. il rate la balle), son adversaire gagne un point. Dans le cas contraire, il remporte un point supplémentaire.

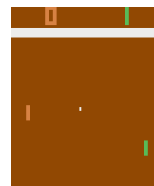


FIGURE 1 – Interface OpenAI Gym pour le jeu Atari Pong

## Description des algorithmes

### Canonical ES

Canonical ES est un algorithme évolutionnaire<sup>1</sup> basique ayant été implémenté par les auteurs de l'article étudié. La démarche est la suivante :

- On tire aléatoirement un vecteur de paramètres  $\theta$
- On génère  $\lambda$  nouveaux candidats (vecteurs de paramètres) en ajoutant du bruit gaussien à  $\theta$
- On évalue chaque candidat en faisant tourner la politique correspondante sur plusieurs épisodes puis on somme les récompenses (scores) qu'il est parvenu à obtenir
- On met ensuite à jour  $\theta$  en lui affectant la moyenne pondérée des  $\mu$  meilleurs candidats puis on recommence

### OpenAI ES

OpenAI ES est l'algorithme évolutionnaire implémenté par les chercheurs d'OpenAI. L'idée est la même que pour Canonical ES sauf que, dans la dernière étape,  $\theta$  devient égal à la somme pondérée de l'ensemble des candidats générés (les coefficients de  $\theta$  deviennent alors proportionnels aux récompenses obtenues par les différents candidats i.e plus le candidat aura obtenu un score élevé plus il aura un poids important).

### CEM

L'algorithme CEM (Cross-Entropy Method) est un algorithme se basant sur une optimisation probabiliste appartenant au domaine de l'optimisation stochastique. Il comprend les deux étapes suivantes :

---

1. Les algorithmes évolutionnaires sont une famille d'algorithmes permettant de résoudre divers problèmes (d'optimisation notamment) en s'inspirant de la théorie de l'évolution. En effet, le principe est de faire évoluer un ensemble de solutions (appelées individus ou candidats) à un problème donné dans le but de trouver les meilleures.

- Générer des échantillons de données aléatoires en utilisant un ensemble de paramètres dynamiques
- Mettre à jour les paramètres régissant la génération de données aléatoires en utilisant les échantillons de données eux-mêmes dans le but d'améliorer les futurs échantillons

## CMA-ES

CMA-ES (Covariance Matrix Adaptation Evolution Strategy) est un algorithme évolutionnaire qui, comme son nom l'indique, se base sur une matrice de covariance. Capable d'adapter sa matrice de covariance et sa moyenne à l'aide des informations provenant des meilleures solutions, l'algorithme peut ainsi décider d'élargir son espace de recherche lorsque les meilleures solutions sont éloignées ou de le réduire lorsque celles-ci sont proches.

## DQN

L'algorithme DQN (Deep Q Network) est un algorithme d'apprentissage par renforcement profond qui vise à estimer la qualité d'effectuer une action à partir d'un état donné, à l'aide de réseaux de neurones. Il prend en entrée l'observation courante et retourne les Q-valeurs de toutes les actions. Ces valeurs estiment la qualité du couple état-action, et ce pour chaque action. Nous avons voulu ajouter DQN à l'ensemble d'algorithmes testés à titre de référence car il est l'un des premiers à avoir battu le score d'un joueur humain professionnel aux jeux Atari.

## Résultats

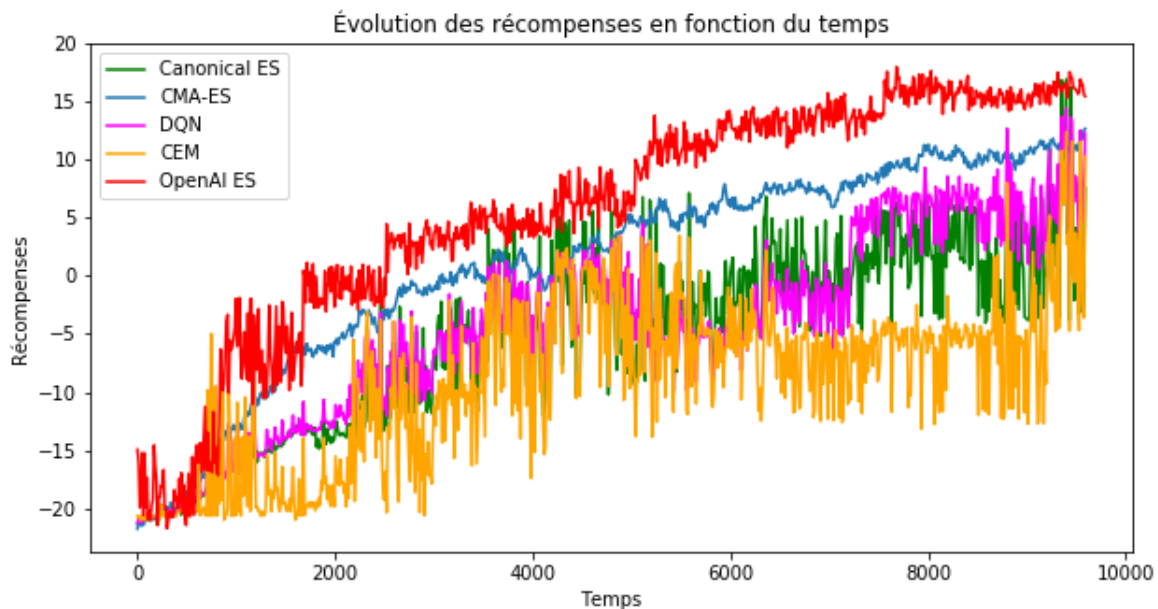


FIGURE 2 – Évolution des récompenses en fonction du temps sur le jeu Pong lors de la phase d'apprentissage

L'entraînement des intelligences artificielles nécessitant d'importantes ressources computationnelles, nous avons opté pour l'utilisation de Google Cloud Platform.

Pour chaque algorithme nous avons effectué nos entraînements sur 10000 itérations puis avons moyenné les récompenses obtenues (la récompense attribuée à l'agent apprenant correspondant à son score lorsque la partie se termine i.e. lui ou son adversaire a atteint 21 points).

On constate qu'OpenAI ES bat tous les autres algorithmes, en particulier Canonical ES. Cependant, bien qu'il soit peu robuste, ce dernier s'est montré assez puissant puisqu'il a pu atteindre les scores enregistrés par OpenAI ES. Les courbes de l'article de base montrent des résultats assez similaires.

L'algorithme CMA-ES s'est également montré très efficace et se place en deuxième position. Vient ensuite DQN, qui lui aussi a pu atteindre des scores assez proches de ceux obtenus par les deux algorithmes précédents. L'algorithme CEM quant à lui se retrouve en dernière position, battu par toutes les méthodes citées précédemment.

On constate que les algorithmes évolutionnaires enregistrent des résultats assez satisfaisants et deviennent de réels concurrents aux algorithmes d'apprentissage par renforcement profond.