

# A Review for Deep Reinforcement Learning in Atari: Benchmarks, Challenges, and Solutions

Anonymous Authors

Anonymous Affiliations  
Anonymous@Anonymous.com

## Abstract

The Arcade Learning Environment (ALE) is proposed as an evaluation platform for empirically assessing the generality of agents across dozens of Atari 2600 games. ALE offers various different and challenging problems, and has drawn great attention from deep reinforcement learning (RL) community. From Deep Q-Networks (DQN) to Agent57, RL agents seems to achieve the superhuman performance in ALE. However, is this really the case? In this paper, we will review the current evaluation metrics in the Atari benchmarks, and then revealed the current judging criteria of achieving superhuman performance is inappropriate, which underestimated the human performance relative to what is possible. To handle those problems and promote the development of RL research, we advocate a novel Atari benchmark based on human world records (HWR), which put forward higher requirement for RL agents on both final performance and learning efficiency. Furthermore, we summarize the state-of-the-art (SOTA) in various problems and provide benchmark results over new evaluation metrics based on human world records. From the benchmark results, we concluded at least four open challenges hindering RL agents to achieve superhuman performance. Finally, we also discuss some promising ways to handle those problems.

## Introduction

The Arcade Learning Environment (Bellemare et al. 2013, ALE) was proposed as a platform for empirically assessing agents designed for general competency across a wide range of Atari games. ALE offers an interface to a diverse set of Atari 2600 game environments designed to be engaging and challenging for human players. As Bellemare et al. (2013) put it, the Atari 2600 games are well suited for evaluating general competency in AI agents for three main reasons:

1. ALE provide multiple different tasks, requiring enough generality.
2. As some of ALE have not been broken through, they are also challenging for humans.
3. Developed by an independent party, ALE are free of experimenter’s bias.

Agents are expected to perform well in as many games as possible **making minimal assumptions about the domain at hand and without the use of game-specific information**. In recent reinforcement learning advances, researchers (Badia et al. 2020a,b; Kapturowski et al. 2018; Ecoffet et al. 2019;

Schrittwieser et al. 2020; Hessel et al. 2021, 2017) are seeking for agents that can achieve superhuman performance. Deep Q-Networks (Mnih et al. 2015, DQN) was the first algorithm to achieve human-level control in a large number of the Atari 2600 games, measured by human normalized scores (HNS). Subsequently, using HNS to assess performance on Atari games has become one of the most widely used benchmarks in deep reinforcement learning (RL). Current state-of-the-art(SOTA) algorithms such as (Badia et al. 2020a, Agent) claimed that they had achieved superhuman performance when they outperform the human baseline uniformly over all Atari 57 games.

It seems that reinforcement learning agents have been able to reach a level beyond humans. But is this really the case? In this paper, we argue that the performance of current reinforcement learning agents is far from the superhuman level from several aspects.

As Toromanoff, Wirbel, and Moutarde (2019) put it, the human baseline scores potentially **underestimating human performance** relative to what is possible. Thus, we argue that this human baseline is far from being representative of the best human player, which means that using it to claim superhuman performance is misleading. In this paper we will propose several challenging problems in the Atari benchmark for superhuman reinforcement learning algorithms.

Learning efficiency (Machado et al. 2018) is one of the indicators to evaluate the learning ability of RL agents. However, many SOTA algorithms (e.g. (Badia et al. 2020a,b; Schrittwieser et al. 2020; Kapturowski et al. 2018)) always overemphasize the final performance they obtain, but ignore the computational cost required to obtain that performance. This has led to more and more algorithms to improve the final performance at the expense of a large number of samples being wasted. In this paper, we argue that an agent that achieve superhuman level not only surpasses human beings in final performance, but also surpasses humans in both learning ability and learning efficiency. In this paper we will propose several measures to to evaluate the learning efficiency of RL agents.

In this work, we first discuss current evaluation metrics of different RL algorithms on ALE and their impact. We then propose a improved evaluation system, extending the recommendations of Toromanoff, Wirbel, and Moutarde (2019); Bellemare et al. (2013); Machado et al. (2018). We suggest

benchmarking on the human world records baseline and show that SOTA RL algorithms are in fact far from superhuman performance on ALE. As an illustration of our new benchmark, several representative algorithms in model-free RL, model-based RL and other fields are benchmarked. Afterwards, we conclude the current existing challenges when we want to design a superhuman agents on Atari and propose some promising solutions.

The main contributions of this work are:

1. **Review of Evaluation Metrics for Atari Benchmark.** We reviewed the mostly used evaluation metrics for ALE and fully discussed the advantages and disadvantages while using those metrics.
2. **Introduction of the measure of learning efficiency for RL agents on ALE.** We argue the importance of learning efficiency for superhuman RL agents and revealed the low learning efficiency problem for current SOTA algorithms (Badia et al. 2020a,b; Kapturowski et al. 2018; Schrittwieser et al. 2020).
3. **Perfection of the world records human baseline.** We provide a complete human world records over all the 57 Atari games, instead part of them (Hafner et al. 2020; Toromanoff, Wirbel, and Moutarde 2019). We further extended the SABER (Toromanoff, Wirbel, and Moutarde 2019) to a more perfect evaluation system with several new evaluation metrics based on human world records.
4. **The proposal, description and justification of an superhuman benchmark for ALE.** We argue the human world records are more representative of the human level than human baselines used in most of previous works.
5. **A novel benchmark results of current state-of-the-art reinforcement learning algorithms.** We review several milestone in the ALE from DQN (Mnih et al. 2015) to GDI (Fan, Xiao, and Huang 2021), and then show the benchmark results of them. with this new benchmark results, we see the ALE is still challenging even for current SOTA algorithms.
6. **Introduction of current challenges and promising solutions for superhuman agents on ALE.** From the new benchmark results, we see massive existing problems before achieving a superhuman agents. We have concluded those problems and provided promising solutions for those problems.

## Background

### Reinforcement Learning

The RL problem can be formulated as a Markov Decision Process (Howard 1960, MDP) defined by  $(\mathcal{S}, \mathcal{A}, p, r, \gamma, \rho_0)$ . Considering a discounted episodic MDP, the initial state  $s_0$  is sampled from the initial distribution  $\rho_0(s) : \mathcal{S} \rightarrow \Delta(\mathcal{S})$ , where we use  $\Delta$  to represent the probability simplex. At each time  $t$ , the agent chooses an action  $a_t \in \mathcal{A}$  according to the policy  $\pi(a_t|s_t) : \mathcal{S} \rightarrow \Delta(\mathcal{A})$  at state  $s_t \in \mathcal{S}$ . The environment receives  $a_t$ , produces the reward  $r_t \sim r(s, a) : \mathcal{S} \times \mathcal{A} \rightarrow \mathbf{R}$  and transfers to the next state  $s_{t+1}$  according to the transition distribution  $p(s' | s, a) : \mathcal{S} \times \mathcal{A} \rightarrow \Delta(\mathcal{S})$ . The process continues until the agent reaches a terminal state or a maximum

time step. Define the discounted state visitation distribution as  $d_{\rho_0}^{\pi}(s) = (1-\gamma)\mathbf{E}_{s_0 \sim \rho_0} [\sum_{t=0}^{\infty} \gamma^t \mathbf{P}(s_t = s | s_0)]$ . The goal of reinforcement learning is to find the optimal policy  $\pi^*$  that maximizes the expected sum of discounted rewards, denoted by  $\mathcal{J}$  (Sutton and Barto 2018):

$$\begin{aligned} \pi^* &= \operatorname{argmax}_{\pi} \mathcal{J}_{\pi} \\ &= \operatorname{argmax}_{\pi} \mathbf{E}_{s_t \sim d_{\rho_0}^{\pi}} \mathbf{E}_{\pi} [G_t | s_t] \\ &= \operatorname{argmax}_{\pi} \mathbf{E}_{s_t \sim d_{\rho_0}^{\pi}} \mathbf{E}_{\pi} \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t \right] \end{aligned} \quad (1)$$

where  $\gamma \in (0, 1)$  is the discount factor.

RL algorithms can be divided into off-policy manners (Mnih et al. 2015, 2016; Haarnoja et al. 2018; Espenholt et al. 2018) and on-policy manners (Schulman et al. 2017). Off-policy algorithms select actions according to a behavior policy  $\mu$  that can be different from the learning policy  $\pi$ . On-policy algorithms evaluate and improve the learning policy through data sampled from the same policy. RL algorithms can also be divided into value-based methods (Mnih et al. 2015; Van Hasselt, Guez, and Silver 2016; Wang et al. 2016; Hessel et al. 2017; Horgan et al. 2018) and policy-based methods (Schulman et al. 2017; Mnih et al. 2016; Espenholt et al. 2018; Schmitt, Hessel, and Simonyan 2020). In the value-based methods, agents learn the policy indirectly, where the policy is defined by consulting the learned value function, like  $\epsilon$ -greedy, and the value function is learned by a typical GPI. In the policy-based methods, agents learn the policy directly, where the correctness of the gradient direction is guaranteed by the policy gradient theorem (Sutton and Barto 2018), and the convergence of the policy gradient methods is also guaranteed (Agarwal et al. 2019).

## Evaluation Metrics for ALE

In this section, we will mainly introduce the evaluation metrics in ALE, including the those have been commonly used by previous works like raw score of all Atari games and the normalized score over all the Atari games based on human average score baseline, and some novel evaluation criteria for the superhuman benchmark including the normalized score based on human world records, learning efficiency, and human world record breakthrough.

### Raw Score

Raw score refers to using tables (e.g. Table of Scores) or figures (e.g. Training Curve) to show the real scores of RL algorithms on all Atari games, which can be calculated by the sum of undiscounted reward of the  $g$ th game of Atari using algorithms  $i$  as follows:

$$G_{g,i} = \mathbf{E}_{s_t \sim d_{\rho_0}^{\pi}} \mathbf{E}_{\pi} \left[ \sum_{k=0}^{\infty} r_{t+k} | s_t \right], g \in [1, 57] \quad (2)$$

As Bellemare et al. (2013) firstly put it, raw score over the whole 57 Atari games can reflect the performance and generality of RL agents to a certain extent. However, this evaluation index has many limitations:

1. It is difficult for to directly compare the performance of the two algorithms from.
2. Its value is easily affected by different game score scales. For example, the scale of Pong is [-21,21], but that of chopper command is [0,999900], so the mean score of those games will simply ignore the performance on Pong.

Therefore, this metric is used to avoid any issues that aggregated metrics may have (Badia et al. 2020a; Fan, Xiao, and Huang 2021). Furthermore, in this paper, we used this metrics to provide evidence that a RL agent surpass the human world record, which will be introduced in detail later.

### Normalized Scores

In order to solve the problem of raw score, some researchers (Bellemare et al. 2013; Mnih et al. 2015) proposed the normalized score to solve those two mentioned problems. The normalized score of the the  $g$ th game of Atari using algorithms  $i$  is calculated as follows:

$$Z_{g,i} = \frac{G_{g,i} - G_{g,base}}{G_{g,reference} - G_{g,base}} \quad (3)$$

As Bellemare et al. (2013) put it, we can compare games with different scoring scales by normalizing scores so that the numerical values become comparable. In practice, we can make  $G_{g,base} = r_{g,min}$  and  $G_{g,reference} = r_{g,max}$ , where  $[r_{g,min}, r_{g,max}]$  is the score scale of the  $g$ th game. And Equ. (3) becomes  $Z_{g,i} = \frac{G_{g,i} - r_{g,min}}{r_{g,max} - r_{g,min}}$ , which is a **Min-Max Scaling** and thus  $Z_{g,i} \in [0, 1]$  become comparable across the 57 games. It seems this metrics can served to compare the performance between two different algorithms. However, the Min-Max normalized score can not intuitively reflect the gap between the algorithm and the average level of humans. Thus, we need a human baseline normalized score.

**Human Average Score Baseline** As we mentioned above, recent reinforcement learning advances (Badia et al. 2020a,b; Kapturowski et al. 2018; Ecoffet et al. 2019; Schrittwieser et al. 2020; Hessel et al. 2021, 2017) are seeking for agents that can achieve superhuman performance. Thus, we need a metric to intuitively reflect the level of the algorithms compared to human performance. As Bellemare et al. (2013) put it, the Human Normalized Score (HNS) is widely used in the RL research (Machado et al. 2018). HNS can be calculated as follows:

$$HNS_{g,i} = \frac{G_{g,i} - G_{g,random}}{G_{g,human\ average} - G_{g,random}} \quad (4)$$

wherein  $g$  donates the  $g$ th game of Atari,  $i$  represents the algorithm  $i$ ,  $G_{i,human}$  represents the human average score baseline (Toromanoff, Wirbel, and Moutarde 2019), and  $G_{g,random}$  represents means the performance of a random policy. Adopting HNS as an evaluation metric of algorithm performance has the following advantages:

1. **Intuitive comparison with human performance.** As  $HNS_{g,i} \geq 100\%$  means algorithm  $i$  have surpass the human average performance in game  $g$ . Therefore, we can directly use HNS to reflect which games the RL agents have surpassed the average human level.

### 2. Performance across algorithms become comparable.

So as the Max-Min Scaling, the human normalized score can also make two different algorithms comparable. The value of  $HNS_{g,i}$  represents the degree to which algorithm  $i$  surpasses the average level of humans in game  $g$ .

**Mean HNS** represents the mean performance of the algorithms across the 57 Atari games. However, it is susceptible to interference from individual high-scoring games like the hard-exploration problems in Atari (Ecoffet et al. 2019). When refer it as the only evaluation metrics, Go-Explore (Ecoffet et al. 2019) has achieved SOTA compared to Agent57 (Badia et al. 2020a), NGU (Badia et al. 2020b), R2D2 (Kapturowski et al. 2018). However, in fact Go-Explore fails to handle many other games in Atari like demon attack, breakout, boxing, phoenix (Fan, Xiao, and Huang 2021). Additionally, Go-Explore fails to balance the trade-off between exploration and exploitation, which makes it suffer from low sample efficiency, which will be discussed in detail in the following.

**Median HNS** represents the median performance of the algorithms across the 57 Atari games. Massive researchers (Schrittwieser et al. 2020; Hessel et al. 2021) has adopted it as a more reasonable metrics for comparing performance between different algorithms. The median HNS has overcome the interference from individual high-scoring games. However, As far as we can see, there are at least two problems while only refer it as the evaluation metrics. First of all, the median HNS only represent the mediocre performance of an algorithms. How about the top performance? One algorithm (Hessel et al. 2021) can easily achieve high median HNS, but at the same time obtain a poor mean HNS by adjusting the hyperparameters of algorithms for games near the median score. It shows that this metrics can show the generality of the algorithms but fail to reflect the potential of the algorithm. And adopting this metrics will urge us to pursue rather mediocre methods.

In practice, we often use **mean HNS** or **median HNS** to show the final performance or generality of an algorithms. Dispute upon whether the mean value or the median value is more representative to show the generality and performance of the algorithms lasts for several years (Mnih et al. 2015; Hessel et al. 2017; Hafner et al. 2020; Hessel et al. 2021; Bellemare et al. 2013; Machado et al. 2018). To avoid any issues that aggregated metrics may have, **we advocate to calculate both of them in the final results** because they serve the different purpose, and we could not judge any algorithm via single one of them.

**Capped Normalized Score** Capped Normalized Score is also widely used in many reinforcement learning advances (Toromanoff, Wirbel, and Moutarde 2019; Badia et al. 2020a). Among those research, Agent57 (Badia et al. 2020a) adopts the capped human normalized score (CHNS) as a better descriptor for evaluating general performance, which can be calculated as  $CHNS = \max\{\min\{HNS, 1\}, 0\}$ . Agent57 claimed CHNS puts an emphasis in the games that are below the average human performance benchmark, and used CHNS to judge whether a algorithms has surpassed the human performance by judging whether  $CHNS \geq 100\%$ . The mean/median CHNS represents the mean/median complete-

ness of surpassing human performance. As far as we can see, there are several problems while adopting this metrics:

1. CHNS fail to reflect the real performance in a specific games. For example,  $CHNS \geq 100\%$  represents the algorithms surpassed the human performance, but fail to reveal how good it the algorithms on this game. For example, from the view of CHNS, Agent57(Badia et al. 2020a) has achieved SOTA performance across 57 Atari games, but while referring to the mean HNS or median HNS, Agent57 lost to MuZero (Fan, Xiao, and Huang 2021).
2. It’s still controversial that using  $CHNS \geq 100\%$  to represent the superhuman performance(Toromanoff, Wirbel, and Moutarde 2019), and this will be solved in this paper.
3. CHNS ignores the low sample efficiency problem as other metrics using normalized scores. This will be discussed and solved in detail in this paper.

In practice, CHNS can serve as a indicator to reflect whether human performance is broken through in a game by the algorithms. The mean/median CHNS can serve as the metrics to reflect the generality of the algorithms.

**Human World Records Baseline** As (Toromanoff, Wirbel, and Moutarde 2019) put it, the Human Average Score Baseline potentially underestimating human performance relative to what is possible. In order to better reflect the performance of the algorithm compared to the human world record, we introduced the complete human world record baseline extened from (Hafner et al. 2020; Toromanoff, Wirbel, and Moutarde 2019) to normalized raw score, which is called the Human World Records Normalized Score (HWRNS). HWRNS can be calculated as follows:

$$HWRNS_{g,i} = \frac{G_{g,i} - G_{g,random}}{G_{g,human\ world\ records} - G_{g,random}} \quad (5)$$

wherein  $g$  donates the  $g$ th game of Atari,  $i$  represents the algorithm  $i$ ,  $G_{i,human}$  represents the huamn world records breakthrough, and  $G_{g,random}$  represents means the performance of a random policy. Adopting HWRNS as an evaluation metric of algorithm performance has the following advantages:

1. **Intuitive comparison with human world records.** As  $HNS_{g,i} \geq 100\%$  means algorithm  $i$  have surpass the human world records performance in game  $g$ . Therefore, we can directly use HWRNS to reflect which games the RL agents have surpassed the huamn world records, which will be used to calculated the human world records breakthrough in the following.
2. **Performance across algorithms become comparable.** So as the Max-Min Scaling, the HWRNS can also make two different algorithms comparable. The value of  $HWRNS_{g,i}$  represents the degree to which algorithm  $i$  surpasses the human world records in game  $g$ .

**Mean HWRNS** represents the mean performance of the algorithms across the 57 Atari games. Compared to mean HNS, mean HWRNS put forward higher requirements on the algorithm. Poor performance algorithm like SimPLe(Kaiser et al. 2019) will can be directly distinguished from other

algorithms. It requires the algorithms to pursue a better performance across all the game rather than concentrate on one or two of them because breaking through any human world record is a huge milestone which puts forward great challenges to the performance and generality of the algorithm. For example, current model-free SOTA algorithms on HNS is Agent57(Badia et al. 2020a), which only acquires 125.92% mean HWRNS, while GDI-H<sup>3</sup> obtained 154.27% mean HWRNS and thus become the new state-of-the-art.

**Median HWRNS** represents the median performance of the algorithms across the 57 Atari games. Compared to Median HNS, median HWRNS also put forward higher requirements on the algorithm. For example, current SOTA reinforcement learning algorithms like Muzero (Schrittwieser et al. 2020) obtain much higher median HNS over GDI-H<sup>3</sup> but relatively lower median HWRNS.

**Capped HWRNS** Capped HWRNS (also called SABER) is firstly proposed and used by (Toromanoff, Wirbel, and Moutarde 2019), which is calculated by  $SABER = \max\{\min\{HWRNS, 2\}, 0\}$ . SABER also has the same problems as CHNS, we won’t repeat them here. For more detail on SABER, you can see (Toromanoff, Wirbel, and Moutarde 2019).

## Learning Efficiency

As we mentioned above, traditional SOTA algorithms ignore the low learning efficiency problem, which makes the data used for training continuously increasing (from 10B (Kaptrowski et al. 2018) to 100B (Badia et al. 2020a)). Increasing in the training volume hinders the application of reinforcement learning algorithms into the real world. In this paper, we advocate not to improve the final performance via increasing the training volume but to improve the learning efficiency. And we advocate to achieve SOTA within 200M training frames for Atari. To evaluate the learning efficiency of an algorithm, we introduce three promising metrics.

**Training Scale** As one of the normally used metrics to reveal the learning efficiency for machine learning algorithms, training scale can also serve the purpose in RL problem. In ALE, training scale means the training frames used for training. In model-based settings, training frames for world modeling or planning via real world model also need to be counted.

**Game Time** Game time is a unique metric of Atari, which means the real-time gameplay (Machado et al. 2018; Fan, Xiao, and Huang 2021). We can used the following formula to calculate this metric:

$$Game\ Time\ (day) = \frac{Num.Frames}{100000*2*24} \quad (6)$$

For example, 200M training frames equals to 42 days real-time gameplay(Fan, Xiao, and Huang 2021), and 100B training frames equals to 20833 days (57 years) real-time gameplay(Badia et al. 2020a). As far as we know, no Atari human world record was achieved by playing a game continuously for more than 57 years because it’s less than 57 years since the birth of Atari games.

**Learning Efficiency** As we mentioned several times while discussing the drawback of normalized score, learning efficiency has been ignored in massive SOTA algorithms. A great number of SOTA algorithms achieved SOTA through training with huge amounts of data, which may equals to 57 continuously playing for a human. In this paper, we argue it is unreasonable to rely on the increase of data to improve the performance of the algorithm. Thus, we proposed the following metric to evaluate the learning efficiency of an algorithm:

$$\text{Learning Efficiency} = \frac{\text{Related Evaluation Metric}}{\text{Num.Frames}} \quad (7)$$

For example, the learning efficiency of an algorithm over mean HNS is  $\frac{\text{mean HNS}}{\text{Num.Frames}}$ , which means the algorithms that can acquire higher mean HNS using lower training frames are better than those acquiring more training data methods.

## Human World Record Breakthrough

As we mentioned above, we need higher requirement to achieve a superhuman agents in Atari. Therefore, as the CHNS of (Badia et al. 2020a), the Human World Record Breakthrough (HWRB) can serve as the metric to reveal whether the algorithm has achieved the real superhuman performance, which can be calculated by  $\text{HWRB} = \sum_{i=1}^{57} (\text{HWRNS} \geq 1)$ .

## An Atari Benchmark for Superhuman Reinforcement Learning Algorithms

As we have fully discussed the evaluation metrics in ALE. In this section, we mainly introduce the Human World Records Benchmark for Reinforcement Learning on Atari. Firstly, we will discuss some methodological differences in ALE benchmarks found in the literature (Bellemare et al. 2013; Machado et al. 2018; Badia et al. 2020a; Hessel et al. 2017). Then, we will introduce the training and evaluation procedures. In the next section we will report the the benchmark results among representative reinforcement learning algorithms.

## Methodological Differences in ALE Benchmarks

**Episode Termination** In the origin benchmark results of ALE (Bellemare et al. 2013), episodes terminate when the all the lives of the player are lost. But some articles (Mnih et al. 2015; Hessel et al. 2017) will end a training episode after ever loss of life for training while ending a episode after losing all lives for testing. It in fact helps the agent to value their lives more and learn to avoid death. We argue that is kind of game-specific knowledge, which should not be concluded in the benchmark for ALE. As (Machado et al. 2018) put it, *we also advocate to use the game over signal be used for termination*.

**Maximum Episode Length** Several related work (Toromanoff, Wirbel, and Moutarde 2019) also noticed this setting of ALE would affect the results of the algorithms. Maximum Episode Length means the maximum number of frames allowed per episode. This parameter ends the episode after a fixed number of time steps even if the game is not over. In most advance in RL (Badia et al. 2020a; Fan, Xiao, and

Huang 2021; Kapturowski et al. 2018; Badia et al. 2020b), this parameter has been set to 30min (equal to  $1\text{E}+5$  frames), while that in (Machado et al. 2018) is set to 5min. In order to put forward higher requirements on learning efficiency of methods, *we advocate to use 30min as the maximum episode length*, which not only require the agents find a optimal solution of the game, but also require it to acquire the optimal score as soon as possible. We argue that the proposal of no maximum episode length (Toromanoff, Wirbel, and Moutarde 2019) is unreasonable because some games like Kangaroo will never stop due to being tracked in a circle.

**Action Set** In the ALE, there are two set of actions for each game, namely the useful set and full set. Instead of using the useful set consist of 4 actions that have been used in massive works (Mnih et al. 2015; Hessel et al. 2017), *we advocate to use the full set of action which consists of 18 actions*.

## Training and Evaluation Procedures

As recommended by (Machado et al. 2018; Toromanoff, Wirbel, and Moutarde 2019), we adopt the same settings in both training and evaluations, which is more realistic.

**Training Procedures** As we mentioned above, in the training phase, *we advocate to use at most 200M frames and end a episode when all the lives are lost or the episode exceeds 30min*. Inside an episode, agent should select an proper action from the full action set.

**Evaluation Procedures** Except for the same setting as training, in evaluation procedures, we should record the training score by averaging k consecutive episodes across the whole training.

**Reporting performance** As recommended by (Machado et al. 2018), we also advocate to report the training score calculated by averaging k consecutive episodes across the whole training. This gives more information about the stability of the training and removes the statistical bias induced when reporting score of the best policy which is today a common practice (Hessel et al. 2017; Mnih et al. 2015; Badia et al. 2020a,b). *Except for the HNS, we advocate that use evaluation metrics based on human world records like HWRB, HWRNS, SABER should be included in the final performance and at the same time the learning efficiency should also be considered while evaluating whether the RL agents have achieved superhuman performance*.

## Benchmark Results

Since most of the previous work do not experiment on the standard human world records benchmark for reinforcement learning in ALE, we will report the final performance of each algorithms and provide the specific benchmarks settings upon the methodological differences as we discussed in the previous section for a fair comparison.

## Model-Free Reinforcement Learning

**Rainbow** Rainbow (Hessel et al. 2017) is classic value-based RL algorithms among the DQN algorithm family,

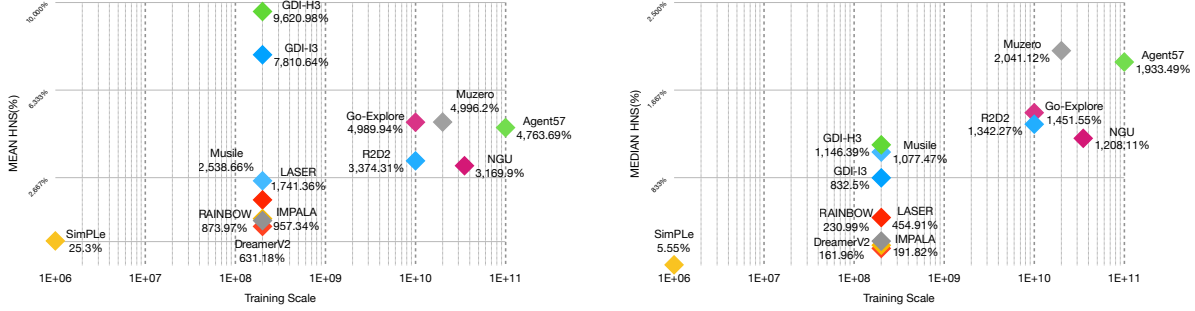


Figure 1: SOTA algorithms of Atari 57 games on mean and median HNS (%) and corresponding training scale.

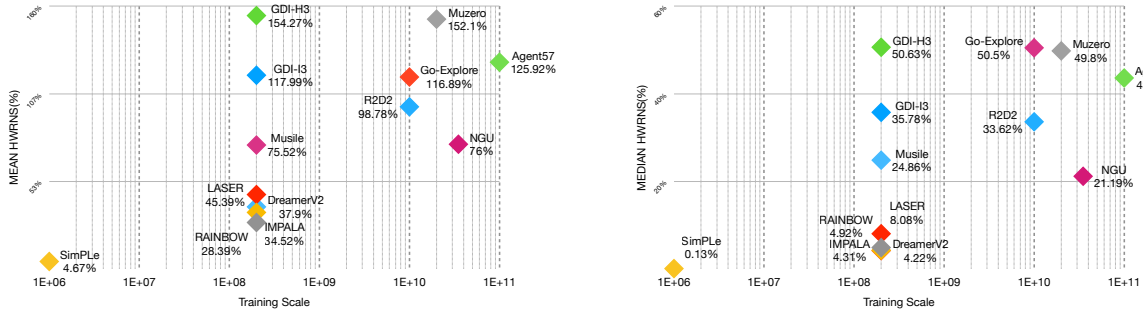


Figure 2: SOTA algorithms of Atari 57 games on mean and median HWRNS (%) and corresponding training scale.

which has fruitfully combined six extensions of DQN algorithm family. It is recognized to achieve the state-of-the-art performance on ALE benchmark. Thus, we select it as one of the representative algorithms of the SOTA DQN algorithms.

**IMPALA** IMPALA, namely the Importance Weighted Actor Learner Architecture, is a classic distributed off-policy actor-critic framework, which decouples acting from learning and learns from experience trajectories using V-trace. IMPALA actors communicate trajectories of experience (sequences of states, actions, and rewards) to a centralized learner which boosts distributed large-scale training. Thus, we select it as one of the representative algorithms of traditional distributed RL algorithm.

**LASER** LASER (Schmitt, Hessel, and Simonyan 2020) is a classic Actor-Critic algorithm, which investigated the combination of Actor-Critic algorithms with a uniform large-scale experience replay. It trained a populations of actors with shared experiences and claimed to achieved SOTA in Atari.

**R2D2** (Kapturowski et al. 2018) Like IMPALA, R2D2 (Kapturowski et al. 2018) is also a classic distributed RL algorithms. It trained RNN-based RL agents from distributed prioritized experience replay, which achieved SOTA in Atari. Thus, we select it as one of the representative value-based distributed RL algorithm.

**NGU** One of the classical problems in ALE for RL agents is the hard exploration problems (Ecoffet et al. 2019; Bellemare et al. 2013; Badia et al. 2020a) like *Private Eye*, *Montezuma’s Revenge*, *Pitfall!*. NGU (Badia et al. 2020b), or Never Give Up, try to ease this problem by augmenting the reward signal with an internally generated intrinsic reward that is sensitive to novelty at two levels: short-term novelty within an episode and long-term novelty across episodes. It then learns a family of policies for exploring and exploiting (sharing the same parameters), with the end goal of obtain the highest score under the exploitative policy. NGU has achieved SOTA in Atari, and thus we selected it as one of the representative population-based model-free RL algorithms.

**Agent57** Agent57 (Badia et al. 2020a) is the SOTA model-free RL algorithms on CHNS or Median HNS of Atari Benchmark. Built on the NGU agents, Agent57 proposed a novel state-action value function parameterization methods and adopted an adaptive exploration over a family of policies, which overcome the drawback of NGU (Badia et al. 2020a). We select it as one of the SOTA model-free RL algorithms.

**GDI** GDI (Fan, Xiao, and Huang 2021), or Generalized Data Distribution Iteration, is the SOTA model-free RL algorithms on mean/median HWRNS, mean HNS, HWRB, median SABER of Atari Benchamrk. GDI is one of the novel Reinforcement Learning paradigms, which combined a data distribution optimization operator into the traditional general-

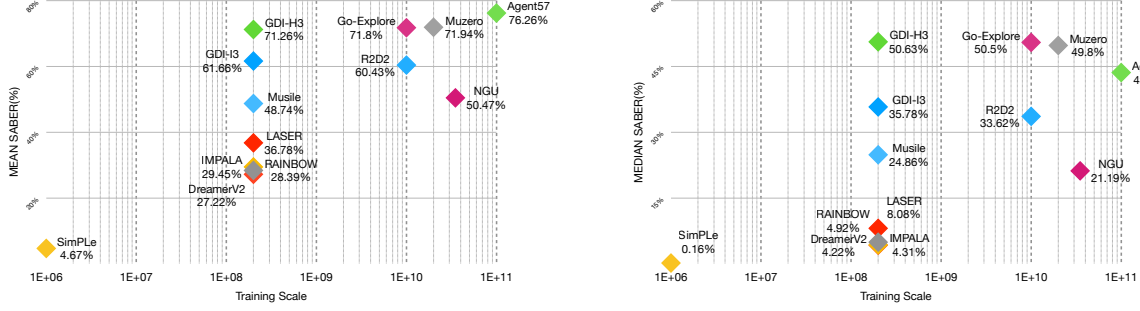


Figure 3: SOTA algorithms of Atari 57 games on mean and median SABER (%) and corresponding training scale.

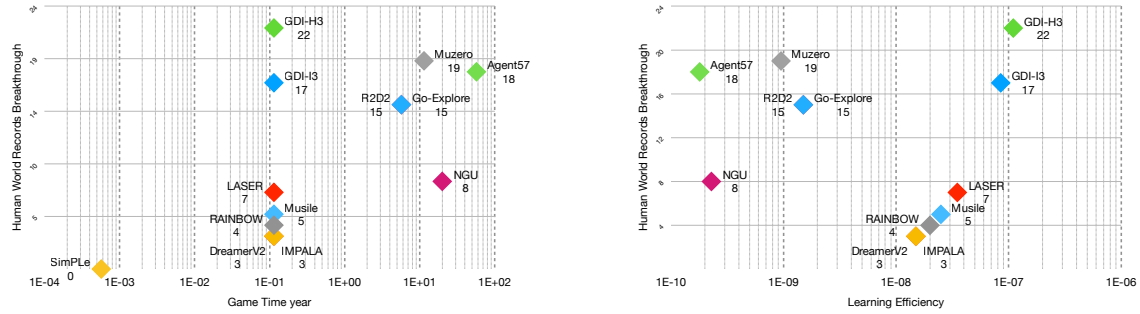


Figure 4: SOTA algorithms of Atari 57 games on HWRB. HWRB of SimPLe is 0, so it’s not shown in the up-right figure.

ized policy iteration (GPI) (Sutton and Barto 2018) and thus achieved human-level learning efficiency. Built on the top of CASA framework (Xiao et al. 2021), GDI-I<sup>3</sup> and GDH<sup>3</sup> have achieved SOTA in Atari Benchmark on Human World Records. Thus, we select them as one of the SOTA model-free RL algorithms.

### Model-Based Reinforcement Learning

**SimPLe** As one of the classic model-based RL algorithms on Atari, SimPLe, or Simulated Policy Learning (Kaiser et al. 2019), adopted a video prediction models to enables RL agents solve Atari problem with higher sample efficiency. It claimed to outperformed the SOTA model-free algorithms in most games, so we select it as one of the representative model-based RL algorithms.

**Dreamer-V2** Dreamer-V2 (Hafner et al. 2020) built world models to facilitate generalization across the past experience and allow learning behaviors from imagined outcomes in the compact latent space of the world model to increase sample efficiency. Dreamer-V2 is claimed to achieve SOTA in Atari, and thus we select it as one of the representative model-based RL algorithms.

**Muzero** Muzero (Schrittwieser et al. 2020) combined a tree-based search with a learned model, and has achieved superhuman performance on Atari. We thus selected it as one of the SOTA model-based RL algorithms.

### Other SOTA algorithms

**Go-Explore** As mentioned in NGU, a grand challenge in reinforcement learning is intelligent exploration, which is called hard-exploration problem (Machado et al. 2018). Go-Explore (Ecoffet et al. 2019) adopted three principles to solve this problem. Firstly, agents remember previously visited states. Secondly, agents first return to a promising state, and then explore from it. Finally, solve simulated environment through any available means, and then robustify via imitation learning. Go-Explore has achieved SOTA in Atari, so we select it as one of the representative SOTA algorithms of hard exploration problem.

**Musile** Musile (Hessel et al. 2021) proposed a novel policy update that combines regularized policy optimization with model learning as an auxiliary loss. It acts directly with a policy network and has computation speed comparable to model-free baselines. As it claimed to achieve SOTA in Atari within 200M training frames, we select it as one of the SOTA sample-efficiency model-based RL algorithms.

### Summary of Benchmark Results

In this part, we summarize the results among all the algorithms we mentioned above on the human world record benchmark for Atari. We illustrated the benchmark results on HNS, HWRNS, SABER and corresponding training scale in Figs. 1, 2 and 3, HWRB and corresponding game time

(year) and learning efficiency in Fig. 4. From those results, we see GDI (Fan, Xiao, and Huang 2021) has achieved SOTA on the benchmark of learning efficiency, HWRB, HWRNS, mean HNS and median SABER within 200M training frames. Agent57 has achieved SOTA in mean SABER and Muzero (Schrittwieser et al. 2020) has achieved SOTA in median HNS. To avoid any issues that aggregated metrics may have, we provide all the raw score of those algorithms in the Appendix.

## Challenges and Solutions

Although RL has achieved fantastic achievements in Atari benchmarks, we could not claim that we have made superhuman agents in Atari. There are still many challenges in the Atari Benchmarks, which also revealed the drawback of current RL algorithms. We believe discuss those challenges and solutions may promote the development of RL research. Therefore, in this section we discuss the current challenges and promising solutions.

### Current Challenges

**Human World Record** From Figs. 4, we see there are at least 35 human world records have not been broken through by current SOTA RL algorithms. Therefore, it’s too early to say we have achieved superhuman performance.

**Hard Exploration Problem** From Tabs 7, 8, and 10 in the appendix, we see current SOTA algorithms are extremely weak facing with the hard exploration problems, and others like Agent57 or Go-Explore that has overcome those problems failed to balance the trade-off between exploration and exploitation leading to a lower learning efficiency and mean HNS.

**Planning and Modeling** Learning from sparse rewards is extremely difficult for model-free RL algorithms, especially for those without an intrinsic rewards which struggles to learn from extremely weak gradient signals. Model-based methods can ease those problems by adopting a world model to planning (Schrittwieser et al. 2020) or replay (Hafner et al. 2020), which both enhanced the gradient signals. However, completely dependent on planning is unrealistic and will lose generality in some Atari Games like Tennis.

**Learning Efficiency** From Figs. 4, current SOTA algorithms like Agent57 may require more than 57 years game-play to achieve SOTA performance, which revealed its low learning efficiency. As recommended in (Hafner et al. 2020), we also argue for high learning efficiency algorithms and we advocate 200M training frames (equal to 41 days) is totally enough for achieving a superhuman agents.

### Promising Solutions

**Adaptive Exploration-Exploitation Balance** The trade-off between exploration-exploitation is a classic difficult problem in RL algorithms (Sutton and Barto 2018). Algorithms designed for hard exploration problems may fail to trade-off the balance leading to low sample efficiency. Others may suffer from hard exploration problems. Thus how to trade off the balance becomes more important. NGU, Agent57 tried to

ease this problems by training a family of policies from extremely explorative to extremely exploitative. Based on that, GDI (Fan, Xiao, and Huang 2021) proposed the data distribution iterator to formulated this procedure, and revealed its superiority to the origin process without the data distribution iterator. It may be a promising way to solve this problem.

**Long Term Planning** Planning algorithms like Muzero (Schrittwieser et al. 2020) fails when the outcome signals of the planning algorithms become misleading or indistinguishable. The former may comes from accumulation of model approximation errors, and the latter may come from a rather sparse rewards environment like Montezuma Revenge. The former problem may need more assistance from more advanced deep learning methods, but the latter can be solved by a longer time planning. More precisely, we need a big picture that guides agents towards a better decision. GDI (Fan, Xiao, and Huang 2021) showed a promising way to combined techniques from Go-Explore and Muzero into the data distribution iteration operator and guide the policy inside an episode, which may solve the low learning efficiency and hard exploration problems. Musile (Hessel et al. 2021) also offer another interesting combination between policy-based methods with model-based RL.

**Compatible Reinforcement Learning Frameworks** As Hessel et al. (2017) put it, DRL community has made several independent improvements those years, but it’s unclear whether there is a unified frameworks that can fruitfully combine those improvements together and make each component compatible. CASA (Xiao et al. 2021) provided a promising frameworks, and based on CASA, GDI further proposed a more general paradigm. We believe a unified frameworks may help to obtain the superhuman agents.

## Conclusion

In this paper, we reviewed the current evaluation metrics for Atari Benchmarks and discussed the advantages and disadvantages of them. We then To further the progress in the field, we proposed the human world records benchmark on Atari, which we suggest for testing a real superhuman agents. Besides, we also provide benchmark results in the human world record benchmark, that may serve as a point of comparison for future work in the ALE. In the final part of this paper, we concluded the challenges and promising solutions that we took from revisiting those milestones in Atari. We also highlighted the current open challenges including: planning and modeling, hard exploration, human world records and low learning efficiency.

## References

- Agarwal, A.; Kakade, S. M.; Lee, J. D.; and Mahajan, G. 2019. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *arXiv preprint arXiv:1908.00261*.
- Badia, A. P.; Piot, B.; Kapturowski, S.; Sprechmann, P.; Vitvitskyi, A.; Guo, D.; and Blundell, C. 2020a. Agent57: Outperforming the atari human benchmark. *arXiv preprint arXiv:2003.13350*.

- Badia, A. P.; Sprechmann, P.; Vitvitskyi, A.; Guo, D.; Piot, B.; Kapturowski, S.; Tieleman, O.; Arjovsky, M.; Pritzel, A.; Bolt, A.; et al. 2020b. Never Give Up: Learning Directed Exploration Strategies. *arXiv preprint arXiv:2002.06038*.
- Bellemare, M. G.; Naddaf, Y.; Veness, J.; and Bowling, M. 2013. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research*, 47: 253–279.
- Ecoffet, A.; Huizinga, J.; Lehman, J.; Stanley, K. O.; and Clune, J. 2019. Go-explore: a new approach for hard-exploration problems. *arXiv preprint arXiv:1901.10995*.
- Espeholt, L.; Soyer, H.; Munos, R.; Simonyan, K.; Mnih, V.; Ward, T.; Doron, Y.; Firoiu, V.; Harley, T.; Dunning, I.; et al. 2018. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. *arXiv preprint arXiv:1802.01561*.
- Fan, J.; Xiao, C.; and Huang, Y. 2021. GDI: Rethinking What Makes Reinforcement Learning Different From Supervised Learning. *arXiv preprint arXiv:2106.06232*.
- Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*.
- Hafner, D.; Lillicrap, T.; Norouzi, M.; and Ba, J. 2020. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*.
- Hessel, M.; Danihelka, I.; Viola, F.; Guez, A.; Schmitt, S.; Sifre, L.; Weber, T.; Silver, D.; and van Hasselt, H. 2021. Muesli: Combining Improvements in Policy Optimization. *arXiv preprint arXiv:2104.06159*.
- Hessel, M.; Modayil, J.; Van Hasselt, H.; Schaul, T.; Ostrovski, G.; Dabney, W.; Horgan, D.; Piot, B.; Azar, M.; and Silver, D. 2017. Rainbow: Combining improvements in deep reinforcement learning. *arXiv preprint arXiv:1710.02298*.
- Horgan, D.; Quan, J.; Budden, D.; Barth-Maron, G.; Hessel, M.; van Hasselt, H.; and Silver, D. 2018. Distributed Prioritized Experience Replay. In *International Conference on Learning Representations*.
- Howard, R. A. 1960. *Dynamic programming and markov processes*. John Wiley.
- Kaiser, L.; Babaeizadeh, M.; Milos, P.; Osinski, B.; Campbell, R. H.; Czechowski, K.; Erhan, D.; Finn, C.; Kozakowski, P.; Levine, S.; et al. 2019. Model-based reinforcement learning for atari. *arXiv preprint arXiv:1903.00374*.
- Kapturowski, S.; Ostrovski, G.; Quan, J.; Munos, R.; and Dabney, W. 2018. Recurrent experience replay in distributed reinforcement learning. In *International conference on learning representations*.
- Machado, M. C.; Bellemare, M. G.; Talvitie, E.; Veness, J.; Hausknecht, M. J.; and Bowling, M. 2018. Revisiting the Arcade Learning Environment: Evaluation Protocols and Open Problems for General Agents. *Journal of Artificial Intelligence Research*, 61: 523–562.
- Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, 1928–1937. PMLR.
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *nature*, 518(7540): 529–533.
- Schmitt, S.; Hessel, M.; and Simonyan, K. 2020. Off-policy actor-critic with shared experience replay. In *International Conference on Machine Learning*, 8545–8554. PMLR.
- Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; et al. 2020. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839): 604–609.
- Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.
- Toromanoff, M.; Wirbel, E.; and Moutarde, F. 2019. Is deep reinforcement learning really superhuman on atari? leveling the playing field. *arXiv preprint arXiv:1908.04683*.
- Van Hasselt, H.; Guez, A.; and Silver, D. 2016. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- Wang, Z.; Schaul, T.; Hessel, M.; Hasselt, H.; Lanctot, M.; and Freitas, N. 2016. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, 1995–2003.
- Xiao, C.; Shi, H.; Fan, J.; and Deng, S. 2021. CASA-B: A Unified Framework of Model-Free Reinforcement Learning. *arXiv:2105.03923*.

# Appendix

## Atari Benchmark Settings

In this part, we will provide the benchmark settings of each algorithms.

| Max episode length Num. Action Repeats Num. Frame Stacks Image Size Grayscaled/RGB Live Information Action Space Dimension |              |          |          |                 |            |           |                  |
|----------------------------------------------------------------------------------------------------------------------------|--------------|----------|----------|-----------------|------------|-----------|------------------|
| RainBow                                                                                                                    | 30min        | 4        | 4        | (84, 84)        | Grayscaled | Yes       | 4                |
| IMPALA                                                                                                                     | 30min        | 4        | 4        | (84, 84)        | Grayscaled | Yes       | 18 (Full)        |
| LASER                                                                                                                      | 30min        | <b>4</b> | <b>4</b> | (84, 84)        | Grayscaled | <b>No</b> | <b>18 (Full)</b> |
| R2D2                                                                                                                       | 30min        | <b>4</b> | 4        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| NGU                                                                                                                        | 30min        | 4        | 1        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| Agent57                                                                                                                    | 30min        | 4        | 1        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| GDI                                                                                                                        | 30min        | 4        | 4        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| SimPLe                                                                                                                     | <b>30min</b> | 4        | 4        | (210, 160)      | RGB        | No        | <b>4</b>         |
| Dreamer-V2                                                                                                                 | 30min        | 4        | 1        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| Muzero                                                                                                                     | 30min        | <b>4</b> | 4        | (84, 84)        | Grayscaled | No        | 18 (Full)        |
| Go-Explore                                                                                                                 | 30min        | 4        | <b>1</b> | <b>(84, 84)</b> | Grayscaled | Yes       | <b>18 (Full)</b> |
| Musile                                                                                                                     | 30min        | 4        | 4        | (96, 96)        | Grayscaled | No        | 18 (Full)        |

Table 1: Atari hyperparameters for training. The values in bold have not been mentioned in the original articles, so we consider them as default values.

| Max episode length Num. Action Repeats Num. Frame Stacks in Image Size Grayscaled/RGB Episode Termination Action Space Dimension Num. Averaging Episodes k |              |          |          |                 |            |                |                  |      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|----------|----------|-----------------|------------|----------------|------------------|------|
| RainBow                                                                                                                                                    | 30min        | 4        | 4        | (84, 84)        | Grayscaled | All lives lost | 4                | 200  |
| IMPALA                                                                                                                                                     | 30min        | 4        | 4        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 200  |
| LASER                                                                                                                                                      | 30min        | <b>4</b> | <b>4</b> | (84, 84)        | Grayscaled | All lives lost | <b>18 (Full)</b> | 100  |
| R2D2                                                                                                                                                       | 30min        | <b>4</b> | 4        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 10   |
| NGU                                                                                                                                                        | 30min        | 4        | 1        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 32   |
| Agent57                                                                                                                                                    | 30min        | 4        | 1        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 50   |
| GDI                                                                                                                                                        | 30min        | 4        | 4        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 32   |
| SimPLe                                                                                                                                                     | <b>30min</b> | 4        | 4        | (210, 160)      | RGB        | All lives lost | <b>4</b>         | 5    |
| Dreamer-V2                                                                                                                                                 | 30min        | 4        | 1        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 10   |
| Muzero                                                                                                                                                     | 30min        | <b>4</b> | 4        | (84, 84)        | Grayscaled | All lives lost | 18 (Full)        | 1000 |
| Go-Explore                                                                                                                                                 | 30min        | 4        | <b>1</b> | <b>(84, 84)</b> | Grayscaled | All lives lost | <b>18 (Full)</b> | 50   |
| Musile                                                                                                                                                     | 30min        | 4        | 4        | (96, 96)        | Grayscaled | All lives lost | 18 (Full)        | 100  |

Table 2: Atari hyperparameters for evaluation. The values in bold have not been mentioned in the original articles, so we consider them as default values.

## Atari Games Table of Scores Based on Human Average Records

In this part, we detail the raw score of several representative SOTA algorithms including the SOTA 200M model-free algorithms, SOTA 10B+ model-free algorithms, SOTA model-based algorithms and other SOTA algorithms.<sup>1</sup> Additionally, we calculate the Human Normalized Score (HNS) of each game with each algorithms. First of all, we demonstrate the sources of the scores that we used. Random scores and average human’s scores are from (Badia et al. 2020a). Rainbow’s scores are from (Hessel et al. 2017). IMPALA’s scores are from (Espeholt et al. 2018). LASER’s scores are from (Schmitt, Hessel, and Simonyan 2020), no sweep at 200M. As there are many versions of R2D2 and NGU, we use original papers’. R2D2’s scores are from (Kapturowski et al. 2018). NGU’s scores are from (Badia et al. 2020b). Agent57’s scores are from (Badia et al. 2020a). MuZero’s scores are from (Schrittwieser et al. 2020). DreamerV2’s scores are from (Hafner et al. 2020). SimPLe’s scores are form (Kaiser et al. 2019). Go-Explore’s scores are form (Ecoffet et al. 2019). Muesli’s scores are form (Hessel et al. 2021). In the following we detail the raw scores and HNS of each algorithms on 57 Atari games.

---

<sup>1</sup>200M and 10B+ represent the training scale.

| Games               | RND     | HUMAN          | RAINBOW        | HNS(%)        | IMPALA           | HNS(%)         | LASER          | HNS(%)         | GDI-I <sup>3</sup> | HNS(%)          | GDI-H <sup>3</sup> | HNS(%)           |
|---------------------|---------|----------------|----------------|---------------|------------------|----------------|----------------|----------------|--------------------|-----------------|--------------------|------------------|
| Scale               |         |                | 200M           |               | 200M             |                | 200M           |                | 200M               |                 | 200M               |                  |
| alien               | 227.8   | 7127.8         | 9491.7         | 134.26        | 15962.1          | 228.03         | 35565.9        | 512.15         | 43384              | 625.45          | <b>48735</b>       | <b>703.00</b>    |
| amidar              | 5.8     | 1719.5         | <b>5131.2</b>  | <b>299.08</b> | 1554.79          | 90.39          | 1829.2         | 106.4          | 1442               | 83.81           | 1065               | 61.81            |
| assault             | 222.4   | 742            | 14198.5        | 2689.78       | 19148.47         | 3642.43        | 21560.4        | 4106.62        | 63876              | 12250.50        | <b>97155</b>       | <b>18655.23</b>  |
| asterix             | 210     | 8503.3         | 428200         | 5160.67       | 300732           | 3623.67        | 240090         | 2892.46        | 759910             | 9160.41         | <b>999999</b>      | <b>12055.38</b>  |
| asteroids           | 719     | 47388.7        | 2712.8         | 4.27          | 108590.05        | 231.14         | 213025         | 454.91         | 751970             | 1609.72         | <b>760005</b>      | <b>1626.94</b>   |
| atlantis            | 12850   | 29028.1        | 826660         | 5030.32       | 849967.5         | 5174.39        | 841200         | 5120.19        | 3803000            | 23427.66        | <b>3837300</b>     | <b>23639.67</b>  |
| bank heist          | 14.2    | 753.1          | 1358           | 181.86        | 1223.15          | 163.61         | 569.4          | 75.14          | <b>1401</b>        | <b>187.68</b>   | 1380               | 184.84           |
| battle zone         | 236     | 37187.5        | 62010          | 167.18        | 20885            | 55.88          | 64953.3        | 175.14         | 478830             | 1295.20         | <b>824360</b>      | <b>2230.29</b>   |
| beam rider          | 363.9   | 16926.5        | 16850.2        | 99.54         | 32463.47         | 193.81         | 90881.6        | 546.52         | 162100             | 976.51          | <b>422890</b>      | <b>2551.09</b>   |
| berzerk             | 123.7   | 2630.4         | 2545.6         | 96.62         | 1852.7           | 68.98          | <b>25579.5</b> | <b>1015.51</b> | 7607               | 298.53          | 14649              | 579.46           |
| bowling             | 23.1    | 160.7          | 30             | 5.01          | 59.92            | 26.76          | 48.3           | 18.31          | 201.9              | 129.94          | <b>205.2</b>       | <b>132.34</b>    |
| boxing              | 0.1     | 12.1           | 99.6           | 829.17        | 99.96            | 832.17         | <b>100</b>     | <b>832.5</b>   | <b>100</b>         | <b>832.50</b>   | <b>100</b>         | <b>832.50</b>    |
| breakout            | 1.7     | 30.5           | 417.5          | 1443.75       | 787.34           | 2727.92        | 747.9          | 2590.97        | <b>864</b>         | <b>2994.10</b>  | <b>864</b>         | <b>2994.10</b>   |
| centipede           | 2090.9  | 12017          | 8167.3         | 61.22         | 11049.75         | 90.26          | <b>292792</b>  | <b>2928.65</b> | 155830             | 1548.84         | 195630             | 1949.80          |
| chopper command     | 811     | 7387.8         | 16654          | 240.89        | 28255            | 417.29         | 761699         | 11569.27       | <b>999999</b>      | <b>15192.62</b> | <b>999999</b>      | <b>15192.62</b>  |
| crazy climber       | 10780.5 | 36829.4        | 168788.5       | 630.80        | 136950           | 503.69         | 167820         | 626.93         | 201000             | 759.39          | <b>241170</b>      | <b>919.76</b>    |
| defender            | 2874.5  | 18688.9        | 55105          | 330.27        | 185203           | 1152.93        | 336953         | 2112.50        | 893110             | 5629.27         | <b>970540</b>      | <b>6118.89</b>   |
| demon attack        | 152.1   | 1971           | 111185         | 6104.40       | 132826.98        | 7294.24        | 133530         | 7332.89        | 675530             | 37131.12        | <b>787985</b>      | <b>43313.70</b>  |
| double dunk         | -18.6   | -16.4          | -0.3           | 831.82        | -0.33            | 830.45         | 14             | 1481.82        | <b>24</b>          | <b>1936.36</b>  | <b>24</b>          | <b>1936.36</b>   |
| enduro              | 0       | 860.5          | 2125.9         | 247.05        | 0                | 0.00           | 0              | 0.00           | <b>14330</b>       | <b>1665.31</b>  | 14300              | 1661.82          |
| fishing derby       | -91.7   | -38.8          | 31.3           | 232.51        | 44.85            | 258.13         | 45.2           | 258.79         | 59                 | 285.71          | <b>65</b>          | <b>296.22</b>    |
| freeway             | 0       | 29.6           | <b>34</b>      | <b>114.86</b> | 0                | 0.00           | 0              | 0.00           | <b>34</b>          | <b>114.86</b>   | <b>34</b>          | <b>114.86</b>    |
| frostbite           | 65.2    | 4334.7         | 9590.5         | 223.10        | 317.75           | 5.92           | 5083.5         | 117.54         | 10485              | 244.05          | <b>11330</b>       | <b>263.84</b>    |
| gopher              | 257.6   | 2412.5         | 70354.6        | 3252.91       | 66782.3          | 3087.14        | 114820.7       | 5316.40        | <b>488830</b>      | <b>22672.63</b> | 473560             | 21964.01         |
| gravitar            | 173     | 3351.4         | 1419.3         | 39.21         | 359.5            | 5.87           | 1106.2         | 29.36          | 5905               | 180.34          | <b>5915</b>        | <b>180.66</b>    |
| hero                | 1027    | 30826.4        | <b>55887.4</b> | <b>184.10</b> | 33730.55         | 109.75         | 31628.7        | 102.69         | 38330              | 125.18          | 38225              | 124.83           |
| ice hockey          | -11.2   | 0.9            | 1.1            | 101.65        | 3.48             | 121.32         | 17.4           | 236.36         | 44.94              | 463.97          | <b>47.11</b>       | <b>481.90</b>    |
| jamesbond           | 29      | 302.8          | 19809          | 72.24         | 601.5            | 209.09         | 37999.8        | 13868.08       | 594500             | 217118.70       | <b>620780</b>      | <b>226716.95</b> |
| kangaroo            | 52      | 3035           | <b>14637.5</b> | <b>488.05</b> | 1632             | 52.97          | 14308          | 477.91         | 14500              | 484.34          | 14636              | 488.00           |
| krull               | 1598    | 2665.5         | 8741.5         | 669.18        | 8147.4           | 613.53         | 9387.5         | 729.70         | 97575              | 8990.82         | <b>594540</b>      | <b>55544.92</b>  |
| kung fu master      | 258.5   | 22736.3        | 52181          | 230.99        | 43375.5          | 191.82         | 607443         | 2701.26        | 140440             | 623.64          | <b>1666665</b>     | <b>7413.57</b>   |
| montezuma revenge   | 0       | <b>4753.3</b>  | 384            | 8.08          | 0                | 0.00           | 0.3            | 0.01           | 3000               | 63.11           | 2500               | 52.60            |
| ms pacman           | 307.3   | 6951.6         | 5380.4         | 76.35         | 7342.32          | 105.88         | 6565.5         | 94.19          | 11536              | 169.00          | <b>11573</b>       | <b>169.55</b>    |
| name this game      | 2292.3  | 8049           | 13136          | 188.37        | 21537.2          | 334.30         | 26219.5        | 415.64         | 34434              | 558.34          | <b>36296</b>       | <b>590.68</b>    |
| phoenix             | 761.5   | 7242.6         | 108529         | 1662.80       | 210996.45        | 3243.82        | 519304         | 8000.84        | 894460             | 13789.30        | <b>959580</b>      | <b>14794.07</b>  |
| pitfall             | -229.4  | <b>6463.7</b>  | 0              | 3.43          | -1.66            | 3.40           | -0.6           | 3.42           | 0                  | 3.43            | -4.345             | 3.36             |
| pong                | -20.7   | 14.6           | 20.9           | 117.85        | 20.98            | 118.07         | <b>21</b>      | <b>118.13</b>  | <b>21</b>          | <b>118.13</b>   | <b>21</b>          | <b>118.13</b>    |
| private eye         | 24.9    | <b>69571.3</b> | 4234           | 6.05          | 98.5             | 0.11           | 96.3           | 0.10           | 15100              | 21.68           | 15100              | 21.68            |
| qbert               | 163.9   | 13455.0        | 33817.5        | 253.20        | <b>351200.12</b> | <b>2641.14</b> | 21449.6        | 160.15         | 27800              | 207.93          | 28657              | 214.38           |
| riverraid           | 1338.5  | 17118.0        | 22920.8        | 136.77        | 29608.05         | 179.15         | <b>40362.7</b> | <b>247.31</b>  | 28075              | 169.44          | 28349              | 171.17           |
| road runner         | 11.5    | 7845           | 62041          | 791.85        | 57121            | 729.04         | 45289          | 578.00         | 878600             | 11215.78        | <b>999999</b>      | <b>12765.53</b>  |
| robotank            | 2.2     | 11.9           | 61.4           | 610.31        | 12.96            | 110.93         | 62.1           | 617.53         | 108.2              | 1092.78         | <b>113.4</b>       | <b>1146.39</b>   |
| seaquest            | 68.4    | 42054.7        | 15898.9        | 37.70         | 1753.2           | 4.01           | 2890.3         | 6.72           | 943910             | 2247.98         | <b>1000000</b>     | <b>2381.57</b>   |
| skiing              | -17098  | <b>-4336.9</b> | -12957.8       | 32.44         | -10180.38        | 54.21          | -29968.4       | -100.86        | -6774              | 80.90           | -6025              | 86.77            |
| solaris             | 1236.3  | <b>12326.7</b> | 3560.3         | 20.96         | 2365             | 10.18          | 2273.5         | 9.35           | 11074              | 88.70           | 9105               | 70.95            |
| space invaders      | 148     | 1668.7         | 18789          | 1225.82       | 43595.78         | 2857.09        | 51037.4        | 3346.45        | 140460             | 9226.80         | <b>154380</b>      | <b>10142.17</b>  |
| star gunner         | 664     | 10250          | 127029         | 1318.22       | 200625           | 2085.97        | 321528         | 3347.21        | 465750             | 4851.72         | <b>677590</b>      | <b>7061.61</b>   |
| surround            | -10     | 6.5            | <b>9.7</b>     | <b>119.39</b> | 7.56             | 106.42         | 8.4            | 111.52         | -7.8               | 13.33           | 2.606              | 76.40            |
| tennis              | -23.8   | -8.3           | 0              | 153.55        | 0.55             | 157.10         | 12.2           | 232.26         | <b>24</b>          | <b>308.39</b>   | <b>24</b>          | <b>308.39</b>    |
| time pilot          | 3568    | 5229.2         | 12926          | 563.36        | 48481.5          | 2703.84        | 105316         | 6125.34        | 216770             | 12834.99        | <b>450810</b>      | <b>26924.45</b>  |
| tutankham           | 11.4    | 167.6          | 241            | 146.99        | 292.11           | 179.71         | 278.9          | 171.25         | <b>423.9</b>       | <b>264.08</b>   | 418.2              | 260.44           |
| up n down           | 533.4   | 11693.2        | 125755         | 1122.08       | 332546.75        | 2975.08        | 345727         | 3093.19        | <b>986440</b>      | <b>8834.45</b>  | 966590             | 8656.58          |
| venture             | 0       | 1187.5         | 5.5            | 0.46          | 0                | 0.00           | 0              | 0.00           | <b>2035</b>        | <b>171.37</b>   | 2000               | 168.42           |
| video pinball       | 0       | 17667.9        | 533936.5       | 3022.07       | 572898.27        | 3242.59        | 511835         | 2896.98        | 925830             | 5240.18         | <b>978190</b>      | <b>5536.54</b>   |
| wizard of wor       | 563.5   | 4756.5         | 17862.5        | 412.57        | 9157.5           | 204.96         | 29059.3        | 679.60         | <b>64239</b>       | <b>1519.90</b>  | 63735              | 1506.59          |
| yars revenge        | 3092.9  | 54576.9        | 102557         | 193.19        | 84231.14         | 157.60         | 166292.3       | 316.99         | <b>972000</b>      | <b>1881.96</b>  | 968090             | 1874.36          |
| zaxxon              | 32.5    | 9173.3         | 22209.5        | 242.62        | 32935.5          | 359.96         | 41118          | 449.47         | 109140             | 1193.63         | <b>216020</b>      | <b>2362.89</b>   |
| MEAN HNS(%)         | 0.00    | 100.00         |                | 873.97        |                  | 957.34         |                | 1741.36        |                    | 7810.6          |                    | <b>9620.98</b>   |
| Learning Efficiency | 0.00    | N/A            |                | 4.37E-08      |                  | 4.79E-08       |                | 8.71E-08       |                    | 3.91E-07        |                    | <b>4.70E-07</b>  |
| MEDIAN HNS(%)       | 0.00    | 100.00         |                | 230.99        |                  | 191.82         |                | 454.91         |                    | 832.5           |                    | <b>1146.39</b>   |
| Learning Efficiency | 0.00    | N/A            |                | 1.15E-08      |                  | 9.59E-09       |                | 2.27E-08       |                    | 4.16E-08        |                    | <b>5.73E-08</b>  |

Table 3: Score table of SOTA 200M model-free algorithms on HNS.

| Games               | R2D2           | HNS(%)         | NGU            | HNS(%)         | AGENT57          | HNS(%)          | GDI-I <sup>3</sup> | HNS(%)          | GDI-H <sup>3</sup> | HNS(%)           |
|---------------------|----------------|----------------|----------------|----------------|------------------|-----------------|--------------------|-----------------|--------------------|------------------|
| Scale               | 10B            |                | 35B            |                | 100B             |                 | 200M               |                 | 200M               |                  |
| alien               | 109038.4       | 1576.97        | 248100         | 3592.35        | <b>297638.17</b> | <b>4310.30</b>  | 43384              | 625.45          | 48735              | 703.00           |
| amidar              | 27751.24       | 1619.04        | 17800          | 1038.35        | <b>29660.08</b>  | <b>1730.42</b>  | 1442               | 83.81           | 1065               | 61.81            |
| assault             | 90526.44       | 17379.53       | 34800          | 6654.66        | 67212.67         | 12892.66        | 63876              | 12250.50        | <b>97155</b>       | <b>18655.23</b>  |
| asterix             | 999080         | 12044.30       | 950700         | 11460.94       | 991384.42        | 11951.51        | 759910             | 9160.41         | <b>999999</b>      | <b>12055.38</b>  |
| asteroids           | 265861.2       | 568.12         | 230500         | 492.36         | 150854.61        | 321.70          | 751970             | 1609.72         | <b>760005</b>      | <b>1626.94</b>   |
| atlantis            | 1576068        | 9662.56        | 1653600        | 10141.80       | 1528841.76       | 9370.64         | 3803000            | 23427.66        | <b>3837300</b>     | <b>23639.67</b>  |
| bank heist          | <b>46285.6</b> | <b>6262.20</b> | 17400          | 2352.93        | 23071.5          | 3120.49         | 1401               | 187.68          | 1380               | 184.84           |
| battle zone         | 513360         | 1388.64        | 691700         | 1871.27        | <b>934134.88</b> | <b>2527.36</b>  | 478830             | 1295.20         | 824360             | 2230.29          |
| beam rider          | 128236.08      | 772.05         | 63600          | 381.80         | 300509.8         | 1812.19         | 162100             | 976.51          | <b>422390</b>      | <b>2548.07</b>   |
| berzerk             | 34134.8        | 1356.81        | 36200          | 1439.19        | <b>61507.83</b>  | <b>2448.80</b>  | 7607               | 298.53          | 14649              | 579.46           |
| bowling             | 196.36         | 125.92         | 211.9          | 137.21         | <b>251.18</b>    | <b>165.76</b>   | 201.9              | 129.94          | 205.2              | 132.34           |
| boxing              | 99.16          | 825.50         | 99.7           | 830.00         | <b>100</b>       | <b>832.50</b>   | <b>100</b>         | <b>832.50</b>   | <b>100</b>         | <b>832.50</b>    |
| breakout            | 795.36         | 2755.76        | 559.2          | 1935.76        | 790.4            | 2738.54         | <b>864</b>         | <b>2994.10</b>  | <b>864</b>         | <b>2994.10</b>   |
| centipede           | 532921.84      | 5347.83        | <b>577800</b>  | <b>5799.95</b> | 412847.86        | 4138.15         | 155830             | 1548.84         | 195630             | 1949.80          |
| chopper command     | 960648         | 14594.29       | 999900         | 15191.11       | 999900           | 15191.11        | <b>999999</b>      | <b>15192.62</b> | <b>999999</b>      | <b>15192.62</b>  |
| crazy climber       | 312768         | 1205.59        | 313400         | 1208.11        | <b>565909.85</b> | <b>2216.18</b>  | 201000             | 759.39          | 241170             | 919.76           |
| defender            | 562106         | 3536.22        | 664100         | 4181.16        | 677642.78        | 4266.80         | 893110             | 5629.27         | <b>970540</b>      | <b>6118.89</b>   |
| demon attack        | 143664.6       | 7890.07        | 143500         | 7881.02        | 143161.44        | 7862.41         | 675530             | 37131.12        | <b>787985</b>      | <b>43313.70</b>  |
| double dunk         | 23.12          | 1896.36        | -14.1          | 204.55         | 23.93            | 1933.18         | <b>24</b>          | <b>1936.36</b>  | <b>24</b>          | <b>1936.36</b>   |
| enduro              | 2376.68        | 276.20         | 2000           | 232.42         | 2367.71          | 275.16          | <b>14330</b>       | <b>1665.31</b>  | 14300              | 1661.82          |
| fishing derby       | 81.96          | 328.28         | 32             | 233.84         | <b>86.97</b>     | <b>337.75</b>   | 59                 | 285.71          | 65                 | 296.22           |
| freeway             | <b>34</b>      | <b>114.86</b>  | 28.5           | 96.28          | 32.59            | 110.10          | <b>34</b>          | <b>114.86</b>   | <b>34</b>          | <b>114.86</b>    |
| frostbite           | 11238.4        | 261.70         | 206400         | 4832.76        | <b>541280.88</b> | <b>12676.32</b> | 10485              | 244.05          | 11330              | 263.84           |
| gopher              | 122196         | 5658.66        | 113400         | 5250.47        | 117777.08        | 5453.59         | <b>488830</b>      | <b>22672.63</b> | 473560             | 21964.01         |
| gravitar            | 6750           | 206.93         | 14200          | 441/32         | <b>19213.96</b>  | <b>599.07</b>   | 5905               | 180.34          | 5915               | 180.66           |
| hero                | 37030.4        | 120.82         | 69400          | 229.44         | <b>114736.26</b> | <b>381.58</b>   | 38330              | 125.18          | 38225              | 124.83           |
| ice hockey          | <b>71.56</b>   | <b>683.97</b>  | -4.1           | 58.68          | 63.64            | 618.51          | 44.94              | 463.97          | 47.11              | 481.90           |
| jamesbond           | 23266          | 8486.85        | 26600          | 9704.53        | 135784.96        | 49582.16        | 594500             | 217118.70       | <b>620780</b>      | <b>226716.95</b> |
| kangaroo            | 14112          | 471.34         | <b>35100</b>   | <b>1174.92</b> | 24034.16         | 803.96          | 14500              | 484.34          | 14636              | 488.90           |
| krull               | 145284.8       | 13460.12       | 127400         | 11784.73       | 251997.31        | 23456.61        | 97575              | 8990.82         | <b>594540</b>      | <b>55544.92</b>  |
| kung fu master      | 200176         | 889.40         | 212100         | 942.45         | 206845.82        | 919.07          | 140440             | 623.64          | <b>1666665</b>     | <b>7413.57</b>   |
| montezuma revenge   | 2504           | 52.68          | <b>10400</b>   | <b>218.80</b>  | 9352.01          | 196.75          | 3000               | 63.11           | 2500               | 52.60            |
| ms pacman           | 29928.2        | 445.81         | 40800          | 609.44         | <b>63994.44</b>  | <b>958.52</b>   | 11536              | 169.00          | 11573              | 169.55           |
| name this game      | 45214.8        | 745.61         | 23900          | 375.35         | <b>54386.77</b>  | <b>904.94</b>   | 34434              | 558.34          | 36296              | 590.68           |
| phoenix             | 811621.6       | 125.11         | 959100         | 14786.66       | 908264.15        | 14002.29        | 894460             | 13789.30        | <b>959580</b>      | <b>14794.07</b>  |
| pitfall             | 0              | 3.43           | 7800           | 119.97         | <b>18756.01</b>  | <b>283.66</b>   | 0                  | 3.43            | -4.3               | 3.36             |
| pong                | <b>21</b>      | <b>118.13</b>  | 19.6           | 114.16         | 20.67            | 117.20          | <b>21</b>          | <b>118.13</b>   | <b>21</b>          | <b>118.13</b>    |
| private eye         | 300            | 0.40           | <b>100000</b>  | <b>143.75</b>  | 79716.46         | 114.59          | 15100              | 21.68           | 15100              | 21.68            |
| qbert               | 161000         | 1210.10        | 451900         | 3398.79        | <b>580328.14</b> | <b>4365.06</b>  | 27800              | 207.93          | 28657              | 214.38           |
| riverraid           | 34076.4        | 207.47         | 36700          | 224.10         | <b>63318.67</b>  | <b>392.79</b>   | 28075              | 169.44          | 28349              | 171.17           |
| road runner         | 498660         | 6365.59        | 128600         | 1641.52        | 243025.8         | 3102.24         | 878600             | 11215.78        | <b>999999</b>      | <b>12765.53</b>  |
| robotank            | <b>132.4</b>   | <b>1342.27</b> | 9.1            | 71.13          | 127.32           | 1289.90         | 108.2              | 1092.78         | 113.4              | 1146.39          |
| sequest             | 999991.84      | 2381.55        | <b>1000000</b> | <b>2381.57</b> | 999997.63        | 2381.56         | 943910             | 2247.98         | <b>1000000</b>     | <b>2381.57</b>   |
| skiing              | -29970.32      | -100.87        | -22977.9       | -46.08         | <b>-4202.6</b>   | <b>101.05</b>   | -6774              | 80.90           | -6025              | 86.77            |
| solaris             | 4198.4         | 26.71          | 4700           | 31.23          | <b>44199.93</b>  | <b>387.39</b>   | 11074              | 88.70           | 9105               | 70.95            |
| space invaders      | 55889          | 3665.48        | 43400          | 2844.22        | 48680.86         | 3191.48         | 140460             | 9226.80         | <b>154380</b>      | <b>10142.17</b>  |
| star gunner         | 521728         | 5435.68        | 414600         | 4318.13        | <b>839573.53</b> | <b>8751.40</b>  | 465750             | 4851.72         | 677590             | 7061.61          |
| surround            | <b>9.96</b>    | <b>120.97</b>  | -9.6           | 2.42           | 9.5              | 118.18          | -7.8               | 13.33           | 2.606              | 76.40            |
| tennis              | <b>24</b>      | <b>308.39</b>  | 10.2           | 219.35         | 23.84            | 307.35          | <b>24</b>          | <b>308.39</b>   | <b>24</b>          | <b>308.39</b>    |
| time pilot          | 348932         | 20791.28       | 344700         | 20536.51       | 405425.31        | 24192.24        | 216770             | 12834.99        | <b>450810</b>      | <b>26924.45</b>  |
| tutankham           | 393.64         | 244.71         | 191.1          | 115.04         | <b>2354.91</b>   | <b>1500.33</b>  | 423.9              | 264.08          | 418.2              | 260.44           |
| up n down           | 542918.8       | 4860.17        | 620100         | 5551.77        | 623805.73        | 5584.98         | <b>986440</b>      | <b>8834.45</b>  | 966590             | 8656.58          |
| venture             | 1992           | 167.75         | 1700           | 143.16         | <b>2623.71</b>   | <b>220.94</b>   | 2035               | 171.37          | 2000               | 168.42           |
| video pinball       | 483569.72      | 2737.00        | 965300         | 5463.58        | <b>992340.74</b> | <b>5616.63</b>  | 925830             | 5240.18         | 978190             | 5536.54          |
| wizard of wor       | 133264         | 3164.81        | 106200         | 2519.35        | <b>157306.41</b> | <b>3738.20</b>  | 64293              | 1519.90         | 63735              | 1506.59          |
| yars revenge        | 918854.32      | 1778.73        | 986000         | 1909.15        | <b>998532.37</b> | <b>1933.49</b>  | 972000             | 1881.96         | 968090             | 1874.36          |
| zaxxon              | 181372         | 1983.85        | 111100         | 1215.07        | <b>249808.9</b>  | <b>2732.54</b>  | 109140             | 1193.63         | 216020             | 2362.89          |
| MEAN HNS(%)         |                | 3374.31        |                | 3169.90        |                  | 4763.69         |                    | 7810.6          |                    | <b>9620.98</b>   |
| Learning Efficiency |                | 3.37E-09       |                | 9.06E-10       |                  | 4.76E-10        |                    | 3.91E-07        |                    | <b>4.81E-07</b>  |
| MEDIAN HNS(%)       |                | 1342.27        |                | 1208.11        |                  | <b>1933.49</b>  |                    | 832.5           |                    | 1146.39          |
| Learning Efficiency |                | 1.34E-09       |                | 3.45E-10       |                  | 1.93E-10        |                    | 4.16E-08        |                    | <b>5.73E-08</b>  |

Table 4: Score table of SOTA model-free algorithms on HNS.

| Games               | MuZero            | HNS(%)          | DreamerV2 | HNS(%)        | SimPLe  | HNS(%)   | GDI-I <sup>3</sup> | HNS(%)          | GDI-H <sup>3</sup> | HNS(%)           |
|---------------------|-------------------|-----------------|-----------|---------------|---------|----------|--------------------|-----------------|--------------------|------------------|
| Scale               | 20B               |                 | 200M      |               | 1M      |          | 200M               |                 | 200M               |                  |
| alien               | <b>741812.63</b>  | <b>10747.61</b> | 3483      | 47.18         | 616.9   | 5.64     | 43384              | 625.45          | 48735              | 703.00           |
| amidar              | <b>28634.39</b>   | <b>1670.57</b>  | 2028      | 118.00        | 74.3    | 4.00     | 1442               | 83.81           | 1065               | 61.81            |
| assault             | <b>143972.03</b>  | <b>27665.44</b> | 7679      | 1435.07       | 527.2   | 58.66    | 63876              | 12250.50        | 97155              | 18655.23         |
| asterix             | 998425            | 12036.40        | 25669     | 306.98        | 1128.3  | 11.07    | 759910             | 9160.41         | <b>999999</b>      | <b>12055.38</b>  |
| asteroids           | 678558.64         | 1452.42         | 3064      | 5.02          | 793.6   | 0.16     | 751970             | 1609.72         | <b>760005</b>      | <b>1626.94</b>   |
| atlantis            | 1674767.2         | 10272.64        | 989207    | 6035.05       | 20992.5 | 50.33    | 3803000            | 23427.66        | <b>3837300</b>     | <b>23639.67</b>  |
| bank heist          | 1278.98           | 171.17          | 1043      | 139.23        | 34.2    | 2.71     | <b>1401</b>        | <b>187.68</b>   | 1380               | 184.84           |
| battle zone         | <b>848623</b>     | <b>2295.95</b>  | 31225     | 83.86         | 4031.2  | 10.27    | 478830             | 1295.20         | 824360             | 2230.29          |
| beam rider          | <b>454993.53</b>  | <b>2744.92</b>  | 12413     | 72.75         | 621.6   | 1.56     | 162100             | 976.51          | 422390             | 2548.07          |
| berzerk             | <b>85932.6</b>    | <b>3423.18</b>  | 751       | 25.02         | N/A     | N/A      | 7607               | 298.53          | 14649              | 579.46           |
| bowling             | <b>260.13</b>     | <b>172.26</b>   | 48        | 18.10         | 30      | 5.01     | 202                | 129.94          | 205.2              | 132.34           |
| boxing              | <b>100</b>        | <b>832.50</b>   | 87        | 724.17        | 7.8     | 64.17    | <b>100</b>         | <b>832.50</b>   | <b>100</b>         | <b>832.50</b>    |
| breakout            | <b>864</b>        | <b>2994.10</b>  | 350       | 1209.38       | 16.4    | 51.04    | <b>864</b>         | <b>2994.10</b>  | <b>864</b>         | <b>2994.10</b>   |
| centipede           | <b>1159049.27</b> | <b>11655.72</b> | 6601      | 45.44         | N/A     | N/A      | 155830             | 1548.84         | 195630             | 1949.80          |
| chopper command     | 991039.7          | 15056.39        | 2833      | 30.74         | 979.4   | 2.56     | <b>999999</b>      | <b>15192.62</b> | <b>999999</b>      | <b>15192.62</b>  |
| crazy climber       | <b>458315.4</b>   | <b>1786.64</b>  | 141424    | 521.55        | 62583.6 | 206.81   | 201000             | 759.39          | 241170             | 919.76           |
| defender            | 839642.95         | 5291.18         | N/A       | N/A           | N/A     | N/A      | 893110             | 5629.27         | <b>970540</b>      | <b>6118.89</b>   |
| demon attack        | 143964.26         | 7906.55         | 2775      | 144.20        | 208.1   | 3.08     | 675530             | 37131.12        | <b>787985</b>      | <b>43313.70</b>  |
| double dunk         | 23.94             | 1933.64         | 22        | 1845.45       | N/A     | N/A      | <b>24</b>          | <b>1936.36</b>  | <b>24</b>          | <b>1936.36</b>   |
| enduro              | 2382.44           | 276.87          | 2112      | 245.44        | N/A     | N/A      | <b>14330</b>       | <b>1665.31</b>  | 14300              | 1661.82          |
| fishing derby       | <b>91.16</b>      | <b>345.67</b>   | 60        | 286.77        | -90.7   | 1.89     | 59                 | 285.71          | 65                 | 296.22           |
| freeway             | 33.03             | 111.59          | <b>34</b> | <b>114.86</b> | 16.7    | 56.42    | <b>34</b>          | <b>114.86</b>   | <b>34</b>          | <b>114.86</b>    |
| frostbite           | <b>631378.53</b>  | <b>14786.59</b> | 15622     | 364.37        | 236.9   | 4.02     | 10485              | 244.05          | 11330              | 263.84           |
| gopher              | 130345.58         | 6036.85         | 53853     | 2487.14       | 596.8   | 15.74    | <b>488830</b>      | <b>22672.6</b>  | 473560             | 21964.01         |
| gravitar            | <b>6682.7</b>     | <b>204.81</b>   | 3554      | 106.37        | 173.4   | 0.01     | 5905               | 180.34          | 5915               | 180.66           |
| hero                | <b>49244.11</b>   | <b>161.81</b>   | 30287     | 98.19         | 2656.6  | 5.47     | 38330              | 125.18          | 38225              | 124.83           |
| ice hockey          | <b>67.04</b>      | <b>646.61</b>   | 29        | 332.23        | -11.6   | -3.31    | 44.94              | 463.97          | 47.11              | 481.90           |
| jamesbond           | 41063.25          | 14986.94        | 9269      | 3374.73       | 100.5   | 26.11    | 594500             | 217118.70       | <b>620780</b>      | <b>226716.95</b> |
| kangaroo            | <b>16763.6</b>    | <b>560.23</b>   | 11819     | 394.47        | 51.2    | -0.03    | 14500              | 484.34          | 14636              | 488.90           |
| krull               | 269358.27         | 25082.93        | 9687      | 757.75        | 2204.8  | 56.84    | 97575              | 8990.82         | <b>594540</b>      | <b>55544.92</b>  |
| kung fu master      | 204824            | 910.08          | 66410     | 294.30        | 14862.5 | 64.97    | 140440             | 623.64          | <b>1666665</b>     | <b>7413.57</b>   |
| montezuma revenge   | 0                 | 0.00            | 1932      | 40.65         | N/A     | N/A      | <b>3000</b>        | <b>63.11</b>    | 2500               | 52.60            |
| ms pacman           | <b>243401.1</b>   | <b>3658.68</b>  | 5651      | 80.43         | 1480    | 17.65    | 11536              | 169.00          | 11573              | 169.55           |
| name this game      | <b>157177.85</b>  | <b>2690.53</b>  | 14472     | 211.57        | 2420.7  | 2.23     | 34434              | 558.34          | 36296              | 590.68           |
| phoenix             | 955137.84         | 14725.53        | 13342     | 194.11        | N/A     | N/A      | 894460             | 13789.30        | <b>959580</b>      | <b>14794.07</b>  |
| pitfall             | <b>0</b>          | <b>3.43</b>     | -1        | 3.41          | N/A     | N/A      | <b>0</b>           | <b>3.43</b>     | -4.3               | 3.36             |
| pong                | <b>21</b>         | <b>118.13</b>   | 19        | 112.46        | 12.8    | 94.90    | <b>21</b>          | <b>118.13</b>   | <b>21</b>          | <b>118.13</b>    |
| private eye         | <b>15299.98</b>   | <b>21.96</b>    | 158       | 0.19          | 35      | 0.01     | 15100              | 21.68           | 15100              | 21.68            |
| qbert               | <b>72276</b>      | <b>542.56</b>   | 162023    | 1217.80       | 1288.8  | 8.46     | 27800              | 207.93          | 28657              | 214.38           |
| riverraid           | <b>323417.18</b>  | <b>2041.12</b>  | 16249     | 94.49         | 1957.8  | 3.92     | 28075              | 169.44          | 28349              | 171.17           |
| road runner         | 613411.8          | 7830.48         | 88772     | 1133.09       | 5640.6  | 71.86    | 878600             | 11215.78        | <b>999999</b>      | <b>12765.53</b>  |
| robotank            | <b>131.13</b>     | <b>1329.18</b>  | 65        | 647.42        | N/A     | N/A      | 108                | 1092.78         | 113.4              | 1146.39          |
| sequest             | 999976.52         | 2381.51         | 45898     | 109.15        | 683.3   | 1.46     | 943910             | 2247.98         | <b>1000000</b>     | <b>2381.57</b>   |
| skiing              | -29968.36         | -100.86         | -8187     | 69.83         | N/A     | N/A      | -6774              | 80.90           | <b>-6025</b>       | <b>86.77</b>     |
| solaris             | 56.62             | -10.64          | 883       | -3.19         | N/A     | N/A      | <b>11074</b>       | <b>88.70</b>    | 9105               | 70.95            |
| space invaders      | 74335.3           | 4878.50         | 2611      | 161.96        | N/A     | N/A      | 140460             | 9226.80         | <b>154380</b>      | <b>10142.17</b>  |
| star gunner         | 549271.7          | 5723.01         | 29219     | 297.88        | N/A     | N/A      | 465750             | 4851.72         | <b>677590</b>      | <b>7061.61</b>   |
| surround            | <b>9.99</b>       | <b>121.15</b>   | N/A       | N/A           | N/A     | N/A      | -7.8               | 13.33           | 2.606              | 76.40            |
| tennis              | 0                 | 153.55          | 23        | 301.94        | N/A     | N/A      | <b>24</b>          | <b>308.39</b>   | <b>24</b>          | <b>308.39</b>    |
| time pilot          | <b>476763.9</b>   | <b>28486.90</b> | 32404     | 1735.96       | N/A     | N/A      | 216770             | 12834.99        | 450810             | 26924.45         |
| tutankham           | <b>491.48</b>     | <b>307.35</b>   | 238       | 145.07        | N/A     | N/A      | 424                | 264.08          | 418.2              | 260.44           |
| up n down           | 715545.61         | 6407.03         | 648363    | 5805.03       | 3350.3  | 25.24    | <b>986440</b>      | <b>8834.45</b>  | 966590             | 8656.58          |
| venture             | 0.4               | 0.03            | 0         | 0.00          | N/A     | N/A      | <b>2035</b>        | <b>171.37</b>   | 2000               | 168.42           |
| video pinball       | <b>981791.88</b>  | <b>5556.92</b>  | 22218     | 125.75        | N/A     | N/A      | 925830             | 5240.18         | 978190             | 5536.54          |
| wizard of wor       | <b>197126</b>     | <b>4687.87</b>  | 14531     | 333.11        | N/A     | N/A      | 64439              | 1523.38         | 63735              | 1506.59          |
| yars revenge        | 553311.46         | 1068.72         | 20089     | 33.01         | 5664.3  | 4.99     | <b>972000</b>      | <b>1881.96</b>  | 968090             | 1874.36          |
| zaxxon              | <b>725853.9</b>   | <b>7940.46</b>  | 18295     | 199.79        | N/A     | N/A      | 109140             | 1193.63         | 216020             | 2362.89          |
| MEAN HNS(%)         |                   | 4996.20         |           | 631.17        |         | 25.3     |                    | 7810.6          |                    | <b>9620.98</b>   |
| Learning Efficiency |                   | 2.50E-09        |           | 3.16E-08      |         | 2.53E-07 |                    | 3.91E-07        |                    | <b>4.81E-07</b>  |
| MEDIAN HNS(%)       |                   | <b>2041.12</b>  |           | 161.96        |         | 5.55     |                    | 832.5           |                    | 1146.39          |
| Learning Efficiency |                   | 1.02E-09        |           | 8.10E-09      |         | 5.55E-08 |                    | 4.16E-08        |                    | <b>5.73E-08</b>  |

Table 5: Score table of SOTA model-based algorithms on HNS.

| Games               | Muesli        | HNS(%)          | Go-Explore    | HNS(%)          | GDI-I <sup>3</sup> | HNS(%)          | GDI-H <sup>3</sup> | HNS(%)           |
|---------------------|---------------|-----------------|---------------|-----------------|--------------------|-----------------|--------------------|------------------|
| Scale               | 200M          |                 | 10B           |                 | 200M               |                 | 200M               |                  |
| alien               | 139409        | 2017.12         | <b>959312</b> | <b>13899.77</b> | 43384              | 625.45          | 48735              | 703.00           |
| amidar              | <b>21653</b>  | <b>1263.18</b>  | 19083         | 1113.22         | 1442               | 83.81           | 1065               | 61.81            |
| assault             | 36963         | 7070.94         | 30773         | 5879.64         | 63876              | 12250.50        | <b>97155</b>       | <b>18655.23</b>  |
| asterix             | 316210        | 3810.30         | 999500        | 12049.37        | 759910             | 9160.41         | <b>999999</b>      | <b>12055.38</b>  |
| asteroids           | 484609        | 1036.84         | 112952        | 240.48          | 751970             | 1609.72         | <b>760005</b>      | <b>1626.94</b>   |
| atlantis            | 1363427       | 8348.18         | 286460        | 1691.24         | 3803000            | 23427.66        | <b>3837300</b>     | <b>23639.67</b>  |
| bank heist          | 1213          | 162.24          | <b>3668</b>   | <b>494.49</b>   | 1401               | 187.68          | 1380               | 184.84           |
| battle zone         | 414107        | 1120.04         | <b>998800</b> | <b>2702.36</b>  | 478830             | 1295.20         | 824360             | 2230.29          |
| beam rider          | 288870        | 1741.91         | 371723        | 2242.15         | 162100             | 976.51          | <b>422390</b>      | <b>2548.07</b>   |
| berzerk             | 44478         | 1769.43         | <b>131417</b> | <b>5237.69</b>  | 7607               | 298.53          | 14649              | 579.46           |
| bowling             | 191           | 122.02          | <b>247</b>    | <b>162.72</b>   | 202                | 129.94          | 205.2              | 132.34           |
| boxing              | 99            | 824.17          | 91            | 757.50          | <b>100</b>         | <b>832.50</b>   | <b>100</b>         | <b>832.50</b>    |
| breakout            | 791           | 2740.63         | 774           | 2681.60         | <b>864</b>         | <b>2994.10</b>  | <b>864</b>         | <b>2994.10</b>   |
| centipede           | <b>869751</b> | <b>8741.20</b>  | 613815        | 6162.78         | 155830             | 1548.84         | 195630             | 1949.80          |
| chopper command     | 101289        | 1527.76         | 996220        | 15135.16        | <b>999999</b>      | <b>15192.62</b> | <b>999999</b>      | <b>15192.62</b>  |
| crazy climber       | 175322        | 656.88          | 235600        | 897.52          | 201000             | 759.39          | <b>241170</b>      | <b>919.76</b>    |
| defender            | 629482        | 3962.26         | N/A           | N/A             | 893110             | 5629.27         | <b>970540</b>      | <b>6118.89</b>   |
| demon attack        | 129544        | 7113.74         | 239895        | 13180.65        | 675530             | 37131.12        | <b>787985</b>      | <b>43313.70</b>  |
| double dunk         | -3            | 709.09          | <b>24</b>     | <b>1936.36</b>  | <b>24</b>          | <b>1936.36</b>  | <b>24</b>          | <b>1936.36</b>   |
| enduro              | 2362          | 274.49          | 1031          | 119.81          | <b>14330</b>       | <b>1665.31</b>  | 14300              | 1661.82          |
| fishing derby       | 51            | 269.75          | <b>67</b>     | <b>300.00</b>   | 59                 | 285.71          | 65                 | 296.22           |
| freeway             | 33            | 111.49          | <b>34</b>     | <b>114.86</b>   | <b>34</b>          | <b>114.86</b>   | <b>34</b>          | <b>114.86</b>    |
| frostbite           | 301694        | 7064.73         | <b>999990</b> | <b>23420.19</b> | 10485              | 244.05          | 11330              | 263.84           |
| gopher              | 104441        | 4834.72         | 134244        | 6217.75         | <b>488830</b>      | <b>22672.63</b> | 473560             | 21964.01         |
| gravitar            | 11660         | 361.41          | <b>13385</b>  | <b>415.68</b>   | 5905               | 180.34          | 5915               | 180.66           |
| hero                | 37161         | 121.26          | 37783         | 123.34          | <b>38330</b>       | <b>125.18</b>   | 38225              | 124.83           |
| ice hockey          | 25            | 299.17          | 33            | 365.29          | 44.94              | 463.97          | <b>47.11</b>       | <b>481.90</b>    |
| jamesbond           | 19319         | 7045.29         | 200810        | 73331.26        | 594500             | 217118.70       | <b>620780</b>      | <b>226716.95</b> |
| kangaroo            | 14096         | 470.80          | <b>24300</b>  | <b>812.87</b>   | 14500              | 484.34          | 14636              | 488.90           |
| krull               | 34221         | 3056.02         | 63149         | 5765.90         | 97575              | 8990.82         | <b>594540</b>      | <b>55544.92</b>  |
| kung fu master      | 134689        | 598.06          | 24320         | 107.05          | 140440             | 623.64          | <b>1666665</b>     | <b>7413.57</b>   |
| montezuma revenge   | 2359          | 49.63           | <b>24758</b>  | <b>520.86</b>   | 3000               | 63.11           | 2500               | 52.60            |
| ms pacman           | 65278         | 977.84          | <b>456123</b> | <b>6860.25</b>  | 11536              | 169.00          | 11573              | 169.55           |
| name this game      | 105043        | 1784.89         | <b>212824</b> | <b>3657.16</b>  | 34434              | 558.34          | 36296              | 590.68           |
| phoenix             | 805305        | 12413.69        | 19200         | 284.50          | 894460             | 13789.30        | <b>959580</b>      | <b>14794.07</b>  |
| pitfall             | 0             | 3.43            | <b>7875</b>   | <b>121.09</b>   | 0                  | 3.43            | -4.3               | 3.36             |
| pong                | 20            | 115.30          | <b>21</b>     | <b>118.13</b>   | <b>21</b>          | <b>118.13</b>   | <b>21</b>          | <b>118.13</b>    |
| private eye         | 10323         | 14.81           | <b>69976</b>  | <b>100.58</b>   | 15100              | 21.68           | 15100              | 21.68            |
| qbert               | 157353        | 1182.66         | <b>999975</b> | <b>7522.41</b>  | 27800              | 207.93          | 28657              | 214.38           |
| riverraid           | <b>47323</b>  | <b>291.42</b>   | 35588         | 217.05          | 28075              | 169.44          | 28349              | 171.17           |
| road runner         | 327025        | 4174.55         | 999900        | 12764.26        | 878600             | 11215.78        | <b>999999</b>      | <b>12765.53</b>  |
| robotank            | 59            | 585.57          | <b>143</b>    | <b>1451.55</b>  | 108                | 1092.78         | 113.4              | 1146.39          |
| seaquest            | 815970        | 1943.26         | 539456        | 1284.68         | 943910             | 2247.98         | <b>1000000</b>     | <b>2381.57</b>   |
| skiing              | -18407        | -10.26          | <b>-4185</b>  | <b>101.19</b>   | -6774              | 80.90           | -6025              | 86.77            |
| solaris             | 3031          | 16.18           | <b>20306</b>  | <b>171.95</b>   | 11074              | 88.70           | 9105               | 70.95            |
| space invaders      | 59602         | 3909.65         | 93147         | 6115.54         | 140460             | 9226.80         | <b>154380</b>      | <b>10142.17</b>  |
| star gunner         | 214383        | 2229.49         | 609580        | 6352.14         | 465750             | 4851.72         | <b>677590</b>      | <b>7061.61</b>   |
| surround            | <b>9</b>      | <b>115.15</b>   | N/A           | N/A             | -8                 | 13.33           | 2.606              | 76.40            |
| tennis              | 12            | 230.97          | <b>24</b>     | <b>308.39</b>   | <b>24</b>          | <b>308.39</b>   | <b>24</b>          | <b>308.39</b>    |
| time pilot          | <b>359105</b> | <b>21403.71</b> | 183620        | 10839.32        | 216770             | 12834.99        | 450810             | 26924.45         |
| tutankham           | 252           | 154.03          | <b>528</b>    | <b>330.73</b>   | 424                | 264.08          | 418.2              | 260.44           |
| up n down           | 649190        | 5812.44         | 553718        | 4956.94         | <b>986440</b>      | <b>8834.45</b>  | 966590             | 8656.58          |
| venture             | 2104          | 177.18          | <b>3074</b>   | <b>258.86</b>   | 2035               | 171.37          | 2000               | 168.42           |
| video pinball       | 685436        | 3879.56         | <b>999999</b> | <b>5659.98</b>  | 925830             | 5240.18         | 978190             | 5536.54          |
| wizard of wor       | 93291         | 2211.48         | <b>199900</b> | <b>4754.03</b>  | 64293              | 1519.90         | 63735              | 1506.59          |
| yars revenge        | 557818        | 1077.47         | <b>999998</b> | <b>1936.34</b>  | 972000             | 1881.96         | 968090             | 1874.36          |
| zaxxon              | 65325         | 714.30          | 18340         | 200.28          | 109140             | 1193.63         | <b>216020</b>      | <b>2362.89</b>   |
| MEAN HNS(%)         |               | 2538.66         |               | 4989.94         |                    | 7810.6          |                    | <b>9620.98</b>   |
| Learning Efficiency |               | 1.27E-07        |               | 4.99E-09        |                    | 3.91E-07        |                    | <b>4.81E-07</b>  |
| MEDIAN HNS(%)       |               | 1077.47         |               | <b>1451.55</b>  |                    | 832.5           |                    | 1146.39          |
| Learning Efficiency |               | 5.39E-08        |               | 1.45E-09        |                    | 4.16E-08        |                    | <b>5.73E-08</b>  |

Table 6: Score table of other SOTA algorithms on HNS.

## Atari Games Table of Scores Based on Human World Records

In this part, we detail the raw score of several representative SOTA algorithms including the SOTA 200M model-free algorithms, SOTA 10B+ model-free algorithms, SOTA model-based algorithms and other SOTA algorithms.<sup>2</sup> Additionally, we calculate the human world records normalized world score (HWRNS) of each game with each algorithms. First of all, we demonstrate the sources of the scores that we used. Random scores are from (Badia et al. 2020a). Human world records (HWR) are from (Hafner et al. 2020; Toromanoff, Wirbel, and Moutarde 2019). Rainbow’s scores are from (Hessel et al. 2017). IMPALA’s scores are from (Espeholt et al. 2018). LASER’s scores are from (Schmitt, Hessel, and Simonyan 2020), no sweep at 200M. As there are many versions of R2D2 and NGU, we use original papers’. R2D2’s scores are from (Kapturowski et al. 2018). NGU’s scores are from (Badia et al. 2020b). Agent57’s scores are from (Badia et al. 2020a). MuZero’s scores are from (Schrittwieser et al. 2020). DreamerV2’s scores are from (Hafner et al. 2020). SimPLe’s scores are from (Kaiser et al. 2019). Go-Explore’s scores are from (Ecoffet et al. 2019). Muesli’s scores are from (Hessel et al. 2021). In the following we detail the raw scores and HWRNS of each algorithms on 57 Atari games.

---

<sup>2</sup>200M and 10B+ represent the training scale.

| Games               | RND     | HWR             | RAINBOW    | HWRNS(%)      | IMPALA    | HWRNS(%) | LASER      | HWRNS(%)      | GDI-I <sup>3</sup> | HWRNS(%)       | GDI-H <sup>3</sup> | HWRNS(%)        |
|---------------------|---------|-----------------|------------|---------------|-----------|----------|------------|---------------|--------------------|----------------|--------------------|-----------------|
| Scale               |         |                 | 200M       |               | 200M      |          | 200M       |               | 200M               |                | 200M               |                 |
| alien               | 227.8   | <b>251916</b>   | 9491.7     | 3.68          | 15962.1   | 6.25     | 976.51     | 14.04         | 43384              | 17.15          | 48735              | 19.27           |
| amidar              | 5.8     | <b>104159</b>   | 5131.2     | 4.92          | 1554.79   | 1.49     | 1829.2     | 1.75          | 1442               | 1.38           | 1065               | 1.02            |
| assault             | 222.4   | 8647            | 14198.5    | 165.90        | 19148.47  | 224.65   | 21560.4    | 253.28        | 63876              | 755.57         | <b>97155</b>       | <b>1150.59</b>  |
| asterix             | 210     | <b>1000000</b>  | 428200     | 42.81         | 300732    | 30.06    | 240090     | 23.99         | 759910             | 75.99          | 999999             | 100.00          |
| asteroids           | 719     | <b>10506650</b> | 2712.8     | 0.02          | 108590.05 | 1.03     | 213025     | 2.02          | 751970             | 7.15           | 760005             | 7.23            |
| atlantis            | 12850   | <b>10604840</b> | 826660     | 7.68          | 849967.5  | 7.90     | 841200     | 7.82          | 3803000            | 35.78          | 3837300            | 36.11           |
| bank heist          | 14.2    | <b>82058</b>    | 1358       | 1.64          | 1223.15   | 1.47     | 569.4      | 0.68          | 1401               | 1.69           | 1380               | 1.66            |
| battle zone         | 236     | 801000          | 62010      | 7.71          | 20885     | 2.58     | 64953.3    | 8.08          | 478830             | 59.77          | <b>824360</b>      | 102.92          |
| beam rider          | 363.9   | <b>999999</b>   | 16850.2    | 1.65          | 32463.47  | 3.21     | 90881.6    | 9.06          | 162100             | 16.18          | 422390             | 42.22           |
| berzerk             | 123.7   | <b>1057940</b>  | 2545.6     | 0.23          | 1852.7    | 0.16     | 25579.5    | 2.41          | 7607               | 0.71           | 14649              | 1.37            |
| bowling             | 23.1    | <b>300</b>      | 30         | 2.49          | 59.92     | 13.30    | 48.3       | 9.10          | 201.9              | 64.57          | 205.2              | 65.76           |
| boxing              | 0.1     | <b>100</b>      | 99.6       | 99.60         | 99.96     | 99.96    | <b>100</b> | <b>100.00</b> | <b>100</b>         | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>   |
| breakout            | 1.7     | <b>864</b>      | 417.5      | 48.22         | 787.34    | 91.11    | 747.9      | 86.54         | <b>864</b>         | <b>100.00</b>  | <b>864</b>         | <b>100.00</b>   |
| centipede           | 2090.9  | <b>1301709</b>  | 8167.3     | 0.47          | 11049.75  | 0.69     | 292792     | 22.37         | 155830             | 11.83          | 195630             | 14.89           |
| chopper command     | 811     | <b>999999</b>   | 16654      | 1.59          | 28255     | 2.75     | 761699     | 76.15         | <b>999999</b>      | <b>100.00</b>  | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 10780.5 | 219900          | 168788.5   | 75.56         | 136950    | 60.33    | 167820     | 75.10         | 201000             | 90.96          | <b>241170</b>      | <b>110.17</b>   |
| defender            | 2874.5  | <b>6010500</b>  | 55105      | 0.87          | 185203    | 3.03     | 336953     | 5.56          | 893110             | 14.82          | 970540             | 16.11           |
| demon attack        | 152.1   | <b>1556345</b>  | 111185     | 7.13          | 132826.98 | 8.53     | 133530     | 8.57          | 675530             | 43.40          | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | -18.6   | 21              | -0.3       | 46.21         | -0.33     | 46.14    | 14         | 82.32         | <b>24</b>          | <b>107.58</b>  | <b>24</b>          | <b>107.58</b>   |
| enduro              | 0       | 9500            | 2125.9     | 22.38         | 0         | 0.00     | 0          | 0.00          | <b>14330</b>       | <b>150.84</b>  | 14300              | 150.53          |
| fishing derby       | -91.7   | <b>71</b>       | 31.3       | 75.60         | 44.85     | 83.93    | 45.2       | 84.14         | 59                 | 92.89          | 65                 | 96.31           |
| freeway             | 0       | <b>38</b>       | 34         | 89.47         | 0         | 0.00     | 0          | 0.00          | 34                 | 89.47          | 34                 | 89.47           |
| frostbite           | 65.2    | <b>454830</b>   | 9590.5     | 2.09          | 317.75    | 0.06     | 5083.5     | 1.10          | 10485              | 2.29           | 11330              | 2.48            |
| gopher              | 257.6   | 355040          | 70354.6    | 19.76         | 66782.3   | 18.75    | 114820.7   | 32.29         | <b>488830</b>      | <b>137.71</b>  | 473560             | 133.41          |
| gravitar            | 173     | <b>162850</b>   | 1419.3     | 0.77          | 359.5     | 0.11     | 1106.2     | 0.57          | 5905               | 3.52           | 5915               | 3.53            |
| hero                | 1027    | <b>1000000</b>  | 55887.4    | 5.49          | 33730.55  | 3.27     | 31628.7    | 3.06          | 38330              | 3.73           | 38225              | 3.72            |
| ice hockey          | -11.2   | 36              | 1.1        | 26.06         | 3.48      | 31.10    | 17.4       | 60.59         | 44.92              | 118.94         | <b>47.11</b>       | <b>123.54</b>   |
| jamesbond           | 29      | 45550           | 19809      | 43.45         | 601.5     | 1.26     | 37999.8    | 83.41         | 594500             | 1305.93        | <b>620780</b>      | <b>1363.66</b>  |
| kangaroo            | 52      | <b>1424600</b>  | 14637.5    | 1.02          | 1632      | 0.11     | 14308      | 1.00          | 14500              | 1.01           | 14636              | 1.02            |
| krull               | 1598    | 104100          | 8741.5     | 6.97          | 8147.4    | 6.39     | 9387.5     | 7.60          | 97575              | 93.63          | <b>594540</b>      | <b>578.47</b>   |
| kung fu master      | 258.5   | 1000000         | 52181      | 5.19          | 43375.5   | 4.31     | 607443     | 60.73         | 140440             | 14.02          | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 0       | <b>1219200</b>  | 384        | 0.03          | 0         | 0.00     | 0.3        | 0.00          | 3000               | 0.25           | 2500               | 0.21            |
| ms pacman           | 307.3   | <b>290090</b>   | 5380.4     | 1.75          | 7342.32   | 2.43     | 6565.5     | 2.16          | 11536              | 3.87           | 11573              | 3.89            |
| name this game      | 2292.3  | 25220           | 13136      | 47.30         | 21537.2   | 83.94    | 26219.5    | 104.36        | 34434              | 140.19         | <b>36296</b>       | <b>148.31</b>   |
| phoenix             | 761.5   | <b>4014440</b>  | 108529     | 2.69          | 210996.45 | 5.24     | 519304     | 12.92         | 894460             | 22.27          | 959580             | 23.89           |
| pitfall             | -229.4  | <b>114000</b>   | 0          | 0.20          | -1.66     | 0.20     | -0.6       | 0.20          | <b>0</b>           | 0.20           | -4.3               | 0.20            |
| pong                | -20.7   | <b>21</b>       | 20.9       | 99.76         | 20.98     | 99.95    | <b>21</b>  | <b>100.00</b> | <b>21</b>          | <b>100.00</b>  | <b>21</b>          | <b>100.00</b>   |
| private eye         | 24.9    | <b>101800</b>   | 4234       | 4.14          | 98.5      | 0.07     | 96.3       | 0.07          | 15100              | 14.81          | 15100              | 14.81           |
| qbert               | 163.9   | <b>2400000</b>  | 33817.5    | 1.40          | 351200.12 | 14.63    | 21449.6    | 0.89          | 27800              | 1.15           | 28657              | 1.19            |
| riverraid           | 1338.5  | <b>1000000</b>  | 22920.8    | 2.16          | 29608.05  | 2.83     | 40362.7    | 3.91          | 28075              | 2.68           | 28349              | 2.70            |
| road runner         | 11.5    | <b>2038100</b>  | 62041      | 3.04          | 57121     | 2.80     | 45289      | 2.22          | 878600             | 43.11          | 999999             | 49.06           |
| robotank            | 2.2     | 76              | 61.4       | 80.22         | 12.96     | 14.58    | 62.1       | 81.17         | 108.2              | 143.63         | <b>113.4</b>       | <b>150.68</b>   |
| sequest             | 68.4    | 999999          | 15898.9    | 1.58          | 1753.2    | 0.17     | 2890.3     | 0.28          | 943910             | 94.39          | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -17098  | <b>-3272</b>    | -12957.8   | 29.95         | -10180.38 | 50.03    | -29968.4   | -93.09        | -6774              | 74.67          | -6025              | 86.77           |
| solaris             | 1236.3  | <b>111420</b>   | 3560.3     | 2.11          | 2365      | 1.02     | 2273.5     | 0.94          | 11074              | 8.93           | 9105               | 7.14            |
| space invaders      | 148     | <b>621535</b>   | 18789      | 3.00          | 43595.78  | 6.99     | 51037.4    | 8.19          | 140460             | 22.58          | 154380             | 24.82           |
| star gunner         | 664     | 77400           | 127029     | 164.67        | 200625    | 260.58   | 321528     | 418.14        | 465750             | 606.09         | <b>677590</b>      | <b>882.15</b>   |
| surround            | -10     | 9.6             | <b>9.7</b> | <b>100.51</b> | 7.56      | 89.59    | 8.4        | 93.88         | -7.8               | 11.22          | 2.606              | 64.32           |
| tennis              | -23.8   | 21              | 0          | 53.13         | 0.55      | 54.35    | 12.2       | 80.36         | <b>24</b>          | <b>106.70</b>  | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 3568    | 65300           | 12926      | 15.16         | 48481.5   | 72.76    | 105316     | 164.82        | 216770             | 345.37         | <b>450810</b>      | <b>724.49</b>   |
| tutankham           | 11.4    | <b>5384</b>     | 241        | 4.27          | 292.11    | 5.22     | 278.9      | 4.98          | 423.9              | 7.68           | 418.2              | 7.57            |
| up n down           | 533.4   | 82840           | 125755     | 152.14        | 332546.75 | 403.39   | 345727     | 419.40        | <b>986440</b>      | <b>1197.85</b> | 966590             | 1173.73         |
| venture             | 0       | <b>38900</b>    | 5.5        | 0.01          | 0         | 0.00     | 0          | 0.00          | 2000               | 5.23           | 2000               | 5.14            |
| video pinball       | 0       | <b>89218328</b> | 533936.5   | 0.60          | 572898.27 | 0.64     | 511835     | 0.57          | 925830             | 1.04           | 978190             | 1.10            |
| wizard of wor       | 563.5   | <b>395300</b>   | 17862.5    | 4.38          | 9157.5    | 2.18     | 29059.3    | 7.22          | 64439              | 16.14          | 63735              | 16.00           |
| yars revenge        | 3092.9  | <b>15000105</b> | 102557     | 0.66          | 84231.14  | 0.54     | 166292.3   | 1.09          | 972000             | 6.46           | 968090             | 6.43            |
| zaxxon              | 32.5    | 83700           | 22209.5    | 26.51         | 32935.5   | 39.33    | 41118      | 49.11         | 109140             | 130.41         | <b>216020</b>      | <b>258.15</b>   |
| MEAN HWRNS(%)       | 0.00    | 100.00          |            | 28.39         |           | 34.52    |            | 45.39         |                    | 117.99         |                    | <b>154.27</b>   |
| Learning Efficiency | 0.00    | N/A             |            | 1.42E-09      |           | 1.73E-09 |            | 2.27E-09      |                    | 5.90E-09       |                    | <b>7.71E-09</b> |
| MEDIAN HWRNS(%)     | 0.00    | <b>100.00</b>   |            | 4.92          |           | 4.31     |            | 8.08          |                    | 35.78          |                    | 50.63           |
| Learning Efficiency | 0.00    | N/A             |            | 2.46E-10      |           | 2.16E-10 |            | 4.04E-10      |                    | 1.79E-09       |                    | <b>2.53E-09</b> |
| HWRB                | 0       | <b>57</b>       |            | 4             |           | 3        |            | 7             |                    | 17             |                    | 22              |

Table 7: Score table of SOTA 200M model-free algorithms on HWRNS

| Games               | R2D2           | HWRNS(%)      | NGU            | HWRNS(%)      | AGENT57          | HWRNS(%)       | GDI-I <sup>3</sup> | HWRNS(%)       | GDI-H <sup>3</sup> | HWRNS(%)        |
|---------------------|----------------|---------------|----------------|---------------|------------------|----------------|--------------------|----------------|--------------------|-----------------|
| Scale               | 10B            |               | 35B            |               | 100B             |                | 200M               |                | 200M               |                 |
| alien               | 109038.4       | 43.23         | 248100         | 98.48         | <b>297638.17</b> | <b>118.17</b>  | 43384              | 17.15          | 48735              | 19.27           |
| amidar              | 27751.24       | 26.64         | 17800          | 17.08         | <b>29660.08</b>  | <b>28.47</b>   | 1442               | 1.38           | 1065               | 1.02            |
| assault             | 90526.44       | 1071.91       | 34800          | 410.44        | 67212.67         | 795.17         | 63876              | 755.57         | <b>97155</b>       | <b>1150.59</b>  |
| asterix             | 999080         | 99.91         | 950700         | 95.07         | 991384.42        | 99.14          | 759910             | 75.99          | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 265861.2       | 2.52          | 230500         | 2.19          | 150854.61        | 1.43           | 751970             | 7.15           | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1576068        | 14.76         | 1653600        | 15.49         | 1528841.76       | 14.31          | 3803000            | 35.78          | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | <b>46285.6</b> | <b>56.40</b>  | 17400          | 21.19         | 23071.5          | 28.10          | 1401               | 1.69           | 1380               | 1.66            |
| battle zone         | 513360         | 64.08         | 691700         | 86.35         | <b>934134.88</b> | <b>116.63</b>  | 478830             | 59.77          | 824360             | 102.92          |
| beam rider          | 128236.08      | 12.79         | 63600          | 6.33          | 300509.8         | 30.03          | 162100             | 16.18          | <b>422390</b>      | <b>42.22</b>    |
| berzerk             | 34134.8        | 3.22          | 36200          | 3.41          | <b>61507.83</b>  | <b>5.80</b>    | 7607               | 0.71           | 14649              | 1.37            |
| bowling             | 196.36         | 62.57         | 211.9          | 68.18         | <b>251.18</b>    | <b>82.37</b>   | 201.9              | 64.57          | 205.2              | 65.76           |
| boxing              | 99.16          | 99.16         | 99.7           | 99.70         | <b>100</b>       | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>   |
| breakout            | 795.36         | 92.04         | 559.2          | 64.65         | 790.4            | 91.46          | <b>864</b>         | <b>100.00</b>  | <b>864</b>         | <b>100.00</b>   |
| centipede           | 532921.84      | 40.85         | <b>577800</b>  | <b>44.30</b>  | 412847.86        | 31.61          | 155830             | 11.83          | 195630             | 14.89           |
| chopper command     | 960648         | 96.06         | 999900         | 99.99         | 999900           | 99.99          | <b>999999</b>      | <b>100.00</b>  | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 312768         | 144.41        | 313400         | 144.71        | <b>565909.85</b> | <b>265.46</b>  | 201000             | 90.96          | 241170             | 110.17          |
| defender            | 562106         | 9.31          | 664100         | 11.01         | 677642.78        | 11.23          | 893110             | 14.82          | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 143664.6       | 9.22          | 143500         | 9.21          | 143161.44        | 9.19           | 675530             | 43.40          | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | 23.12          | 105.35        | -14.1          | 11.36         | 23.93            | 107.40         | <b>24</b>          | <b>107.58</b>  | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2376.68        | 25.02         | 2000           | 21.05         | 2367.71          | 24.92          | <b>14330</b>       | <b>150.84</b>  | 14300              | 150.53          |
| fishing derby       | 81.96          | 106.74        | 32             | 76.03         | <b>86.97</b>     | <b>109.82</b>  | 59                 | 92.89          | 65                 | 96.31           |
| freeway             | <b>34</b>      | <b>89.47</b>  | 28.5           | 75.00         | 32.59            | 85.76          | <b>34</b>          | <b>89.47</b>   | <b>34</b>          | <b>89.47</b>    |
| frostbite           | 11238.4        | 2.46          | 206400         | 45.37         | <b>541280.88</b> | <b>119.01</b>  | 10485              | 2.29           | 11330              | 2.48            |
| gopher              | 122196         | 34.37         | 113400         | 31.89         | 117777.08        | 33.12          | <b>488830</b>      | <b>137.71</b>  | 473560             | 133.41          |
| gravitar            | 6750           | 4.04          | 14200          | 8.62          | <b>19213.96</b>  | <b>11.70</b>   | 5905               | 3.52           | 5915               | 3.53            |
| hero                | 37030.4        | 3.60          | 69400          | 6.84          | <b>114736.26</b> | <b>11.38</b>   | 38330              | 3.73           | 38225              | 3.72            |
| ice hockey          | <b>71.56</b>   | <b>175.34</b> | -4.1           | 15.04         | 63.64            | 158.56         | 37.89              | 118.94         | 47.11              | 123.54          |
| jamesbond           | 23266          | 51.05         | 26600          | 58.37         | 135784.96        | 298.23         | 594500             | 1305.93        | <b>620780</b>      | <b>1363.66</b>  |
| kangaroo            | 14112          | 0.99          | <b>35100</b>   | <b>2.46</b>   | 24034.16         | 1.68           | 14500              | 1.01           | 14636              | 1.02            |
| krull               | 145284.8       | 140.18        | 127400         | 122.73        | 251997.31        | 244.29         | 97575              | 93.63          | <b>594540</b>      | <b>578.47</b>   |
| kung fu master      | 200176         | 20.00         | 212100         | 21.19         | 206845.82        | 20.66          | 140440             | 14.02          | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 2504           | 0.21          | <b>10400</b>   | <b>0.85</b>   | 9352.01          | 0.77           | 3000               | 0.25           | 2500               | 0.21            |
| ms pacman           | 29928.2        | 10.22         | 40800          | 13.97         | <b>63994.44</b>  | <b>21.98</b>   | 11536              | 3.87           | 11573              | 3.89            |
| name this game      | 45214.8        | 187.21        | 23900          | 94.24         | <b>54386.77</b>  | <b>227.21</b>  | 34434              | 140.19         | 36296              | 148.31          |
| phoenix             | 811621.6       | 20.20         | <b>959100</b>  | <b>23.88</b>  | 908264.15        | 22.61          | 894460             | 22.27          | 959580             | 23.89           |
| pitfall             | 0              | 0.20          | 7800           | 7.03          | <b>18756.01</b>  | <b>16.62</b>   | 0                  | 0.20           | -4.3               | 0.20            |
| pong                | <b>21</b>      | <b>100.00</b> | 19.6           | 96.64         | 20.67            | 99.21          | <b>21</b>          | <b>100.00</b>  | <b>21</b>          | <b>100.00</b>   |
| private eye         | 300            | 0.27          | <b>100000</b>  | <b>98.23</b>  | 79716.46         | 78.30          | 15100              | 14.81          | 15100              | 14.81           |
| qbert               | 161000         | 6.70          | 451900         | 18.82         | <b>580328.14</b> | <b>24.18</b>   | 27800              | 1.15           | 28657              | 1.19            |
| riverraid           | 34076.4        | 3.28          | 36700          | 3.54          | <b>63318.67</b>  | <b>6.21</b>    | 28075              | 2.68           | 28349              | 2.70            |
| road runner         | 498660         | 24.47         | 128600         | 6.31          | 243025.8         | 11.92          | 878600             | 43.11          | <b>999999</b>      | <b>49.06</b>    |
| robotank            | <b>132.4</b>   | <b>176.42</b> | 9.1            | 9.35          | 127.32           | 169.54         | 108                | 143.63         | 113.4              | 150.68          |
| seaquest            | 999991.84      | 100.00        | <b>1000000</b> | <b>100.00</b> | 999997.63        | 100.00         | 943910             | 94.39          | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -29970.32      | -93.10        | -22977.9       | -42.53        | <b>-4202.6</b>   | <b>93.27</b>   | -6774              | 74.67          | -6025              | 86.77           |
| solaris             | 4198.4         | 2.69          | 4700           | 3.14          | <b>44199.93</b>  | <b>38.99</b>   | 11074              | 8.93           | 9105               | 7.14            |
| space invaders      | 55889          | 8.97          | 43400          | 6.96          | 48680.86         | 7.81           | 140460             | 22.58          | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 521728         | 679.03        | 414600         | 539.43        | <b>839573.53</b> | <b>1093.24</b> | 465750             | 606.09         | 677590             | 882.15          |
| surround            | <b>9.96</b>    | <b>101.84</b> | -9.6           | 2.04          | 9.5              | 99.49          | -7.8               | 11.22          | 2.606              | 64.32           |
| tennis              | <b>24</b>      | <b>106.70</b> | 10.2           | 75.89         | 23.84            | 106.34         | <b>24</b>          | <b>106.70</b>  | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 348932         | 559.46        | 344700         | 552.60        | <b>405425.31</b> | <b>650.97</b>  | 216770             | 345.37         | 450810             | 724.49          |
| tutankham           | 393.64         | 7.11          | 191.1          | 3.34          | <b>2354.91</b>   | <b>43.62</b>   | 423.9              | 7.68           | 418.2              | 7.57            |
| up n down           | 542918.8       | 658.98        | 620100         | 752.75        | 623805.73        | 757.26         | <b>986440</b>      | <b>1197.85</b> | 966590             | 1173.73         |
| venture             | 1992           | 5.12          | 1700           | 4.37          | <b>2623.71</b>   | <b>6.74</b>    | 2000               | 5.23           | 2000               | 5.14            |
| video pinball       | 483569.72      | 0.54          | 965300         | 1.08          | <b>992340.74</b> | <b>1.11</b>    | 925830             | 1.04           | 978190             | 1.10            |
| wizard of wor       | 133264         | 33.62         | 106200         | 26.76         | <b>157306.41</b> | <b>39.71</b>   | 64439              | 16.14          | 63735              | 16.00           |
| yars revenge        | 918854.32      | 6.11          | 986000         | 6.55          | <b>998532.37</b> | <b>6.64</b>    | 972000             | 6.46           | 968090             | 6.43            |
| zaxxon              | 181372         | 216.74        | 111100         | 132.75        | <b>249808.9</b>  | <b>298.53</b>  | 109140             | 130.41         | 216020             | 258.15          |
| MEAN HWRNS(%)       |                | 98.78         |                | 76.00         |                  | 125.92         |                    | 117.99         |                    | <b>154.27</b>   |
| Learning Efficiency |                | 9.88E-11      |                | 2.17E-11      |                  | 1.26E-11       |                    | 5.90E-09       |                    | <b>7.71E-09</b> |
| MEDIAN HWRNS(%)     |                | 33.62         |                | 21.19         |                  | 43.62          |                    | 35.78          |                    | <b>50.63</b>    |
| Learning Efficiency |                | 3.36E-11      |                | 6.05E-12      |                  | 4.36E-12       |                    | 1.79E-09       |                    | <b>2.53E-09</b> |
| HWRB                |                | 15            |                | 8             |                  | 18             |                    | 17             |                    | <b>22</b>       |

Table 8: Score table of SOTA 10B+ model-free algorithms on HWRNS.

| Games               | MuZero            | HWRNS(%)       | DreamerV2 | HWRNS(%)     | SimPLe  | HWRNS(%)        | GDI-I <sup>3</sup> | HWRNS(%)       | GDI-H <sup>3</sup> | HWRNS(%)        |
|---------------------|-------------------|----------------|-----------|--------------|---------|-----------------|--------------------|----------------|--------------------|-----------------|
| Scale               | 20B               |                | 200M      |              | 1M      |                 | 200M               |                | 200M               |                 |
| alien               | <b>741812.63</b>  | <b>294.64</b>  | 3483      | 1.29         | 616.9   | 0.15            | 43384              | 17.15          | 48735              | 19.27           |
| amidar              | <b>28634.39</b>   | <b>27.49</b>   | 2028      | 1.94         | 74.3    | 0.07            | 1442               | 1.38           | 1065               | 1.02            |
| assault             | <b>143972.03</b>  | <b>1706.31</b> | 7679      | 88.51        | 527.2   | 3.62            | 63876              | 755.57         | 97155              | 1150.59         |
| asterix             | 998425            | 99.84          | 25669     | 2.55         | 1128.3  | 0.09            | 759910             | 75.99          | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 678558.64         | 6.45           | 3064      | 0.02         | 793.6   | 0.00            | 751970             | 7.15           | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1674767.2         | 15.69          | 989207    | 9.22         | 20992.5 | 0.08            | 3803000            | 35.78          | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | 1278.98           | 1.54           | 1043      | 1.25         | 34.2    | 0.02            | <b>1401</b>        | <b>1.69</b>    | 1380               | 1.66            |
| battle zone         | <b>848623</b>     | <b>105.95</b>  | 31225     | 3.87         | 4031.2  | 0.47            | 478830             | 59.77          | 824360             | 102.92          |
| beam rider          | <b>454993.53</b>  | <b>45.48</b>   | 12413     | 1.21         | 621.6   | 0.03            | 162100             | 16.18          | 422390             | 42.22           |
| berzerk             | <b>85932.6</b>    | <b>8.11</b>    | 751       | 0.06         | N/A     | N/A             | 7607               | 0.71           | 14649              | 1.37            |
| bowling             | <b>260.13</b>     | <b>85.60</b>   | 48        | 8.99         | 30      | 2.49            | 202                | 64.57          | 205.2              | 65.76           |
| boxing              | <b>100</b>        | <b>100.00</b>  | 87        | 86.99        | 7.8     | 7.71            | <b>100</b>         | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>   |
| breakout            | <b>864</b>        | <b>100.00</b>  | 350       | 40.39        | 16.4    | 1.70            | <b>864</b>         | <b>100.00</b>  | <b>864</b>         | 100.00          |
| centipede           | <b>1159049.27</b> | <b>89.02</b>   | 6601      | 0.35         | N/A     | N/A             | 155830             | 11.83          | 195630             | 14.89           |
| chopper command     | 991039.7          | 99.10          | 2833      | 0.20         | 979.4   | 0.02            | <b>999999</b>      | <b>100.00</b>  | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | <b>458315.4</b>   | <b>214.01</b>  | 141424    | 62.47        | 62583.6 | 24.77           | 201000             | 90.96          | 241170             | 110.17          |
| defender            | 839642.95         | 13.93          | N/A       | N/A          | N/A     | N/A             | 893110             | 14.82          | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 143964.26         | 9.24           | 2775      | 0.17         | 208.1   | 0.00            | 675530             | 43.40          | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | 23.94             | 107.42         | 22        | 102.53       | N/A     | N/A             | <b>24</b>          | <b>107.58</b>  | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2382.44           | 25.08          | 2112      | 22.23        | N/A     | N/A             | <b>14330</b>       | <b>150.84</b>  | 14300              | 150.53          |
| fishing derby       | <b>91.16</b>      | <b>112.39</b>  | 93.24     | 286.77       | -90.7   | 0.61            | 59                 | 92.89          | 65                 | 96.31           |
| freeway             | 33.03             | 86.92          | <b>34</b> | <b>89.47</b> | 16.7    | 43.95           | <b>34</b>          | <b>89.47</b>   | <b>34</b>          | <b>89.47</b>    |
| frostbite           | <b>631378.53</b>  | <b>138.82</b>  | 15622     | 3.42         | 236.9   | 0.04            | 10485              | 2.29           | 11330              | 2.48            |
| gopher              | 130345.58         | 36.67          | 53853     | 15.11        | 596.8   | 0.10            | <b>488830</b>      | <b>137.71</b>  | 473560             | 133.41          |
| gravitar            | <b>6682.7</b>     | <b>4.00</b>    | 3554      | 2.08         | 173.4   | 0.00            | 5905               | 3.52           | 5915               | 3.53            |
| hero                | <b>49244.11</b>   | <b>4.83</b>    | 30287     | 2.93         | 2656.6  | 0.16            | 38330              | 3.73           | 38225              | 3.72            |
| ice hockey          | <b>67.04</b>      | <b>165.76</b>  | 29        | 85.17        | -11.6   | -0.85           | 38                 | 118.94         | 47.11              | 123.54          |
| jamesbond           | 41063.25          | 90.14          | 9269      | 20.30        | 100.5   | 0.16            | 594500             | 1305.93        | <b>620780</b>      | <b>1363.66</b>  |
| kangaroo            | <b>16763.6</b>    | <b>1.17</b>    | 11819     | 0.83         | 51.2    | 0.00            | 14500              | 1.01           | 14636              | 1.02            |
| krull               | 269358.27         | 261.22         | 9687      | 7.89         | 2204.8  | 0.59            | 97575              | 93.63          | <b>594540</b>      | <b>578.47</b>   |
| kung fu master      | 204824            | 20.46          | 66410     | 6.62         | 14862.5 | 1.46            | 140440             | 14.02          | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 0                 | 0.00           | 1932      | 0.16         | N/A     | N/A             | <b>3000</b>        | <b>0.25</b>    | 2500               | 0.21            |
| ms pacman           | <b>243401.1</b>   | <b>83.89</b>   | 5651      | 1.84         | 1480    | 0.40            | 11536              | 3.87           | 11573              | 3.89            |
| name this game      | <b>157177.85</b>  | <b>675.54</b>  | 14472     | 53.12        | 2420.7  | 0.56            | 34434              | 140.19         | 36296              | 148.31          |
| phoenix             | <b>955137.84</b>  | <b>23.78</b>   | 13342     | 0.31         | N/A     | N/A             | 894460             | 22.27          | 959580             | 23.89           |
| pitfall             | <b>0</b>          | <b>0.20</b>    | -1        | 0.20         | N/A     | N/A             | <b>0</b>           | <b>0.20</b>    | -4.3               | 0.20            |
| pong                | <b>21</b>         | 100.00         | 19        | 95.20        | 12.8    | 80.34           | <b>21</b>          | <b>100.00</b>  | <b>21</b>          | <b>100.00</b>   |
| private eye         | <b>15299.98</b>   | <b>15.01</b>   | 158       | 0.13         | 35      | 0.01            | 15100              | 14.81          | 15100              | 14.81           |
| qbert               | <b>72276</b>      | <b>3.00</b>    | 162023    | 6.74         | 1288.8  | 0.05            | 27800              | 1.15           | 28657              | 1.19            |
| riverraid           | <b>323417.18</b>  | <b>32.25</b>   | 16249     | 1.49         | 1957.8  | 0.06            | 28075              | 2.68           | 28349              | 2.70            |
| road runner         | 613411.8          | 30.10          | 88772     | 4.36         | 5640.6  | 0.28            | 878600             | 43.11          | <b>999999</b>      | <b>49.06</b>    |
| robotank            | <b>131.13</b>     | <b>174.70</b>  | 65        | 85.09        | N/A     | N/A             | 108                | 143.63         | 113.4              | 150.68          |
| seaquest            | 999976.52         | 100.00         | 45898     | 4.58         | 683.3   | 0.06            | 943910             | 94.39          | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -29968.36         | -93.09         | -8187     | 64.45        | N/A     | N/A             | -6774              | 74.67          | <b>-6025</b>       | <b>86.77</b>    |
| solaris             | 56.62             | -1.07          | 883       | -0.32        | N/A     | N/A             | <b>11074</b>       | <b>8.93</b>    | 9105               | 7.14            |
| space invaders      | 74335.3           | 11.94          | 2611      | 0.40         | N/A     | N/A             | 140460             | 22.58          | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 549271.7          | 714.93         | 29219     | 37.21        | N/A     | N/A             | 465750             | 606.09         | <b>677590</b>      | <b>882.15</b>   |
| surround            | <b>9.99</b>       | <b>101.99</b>  | N/A       | N/A          | N/A     | N/A             | -8                 | 11.22          | 2.606              | 64.32           |
| tennis              | 0                 | 53.13          | 23        | 104.46       | N/A     | N/A             | <b>24</b>          | <b>106.70</b>  | <b>24</b>          | <b>106.70</b>   |
| time pilot          | <b>476763.9</b>   | <b>766.53</b>  | 32404     | 46.71        | N/A     | N/A             | 216770             | 345.37         | 450810             | 724.49          |
| tutankham           | <b>491.48</b>     | <b>8.94</b>    | 238       | 4.22         | N/A     | N/A             | 424                | 7.68           | 418.2              | 7.57            |
| up n down           | 715545.61         | 868.72         | 648363    | 787.09       | 3350.3  | 3.42            | <b>986440</b>      | <b>1197.85</b> | 966590             | 1173.73         |
| venture             | 0.4               | 0.00           | 0         | 0.00         | N/A     | N/A             | <b>2030</b>        | <b>5.23</b>    | 2000               | 5.14            |
| video pinball       | <b>981791.88</b>  | <b>1.10</b>    | 22218     | 0.02         | N/A     | N/A             | 925830             | 1.04           | 978190             | 1.10            |
| wizard of wor       | <b>197126</b>     | <b>49.80</b>   | 14531     | 3.54         | N/A     | N/A             | 64439              | 16.14          | 63735              | 16.00           |
| yars revenge        | 553311.46         | 3.67           | 20089     | 0.11         | 5664.3  | 0.02            | <b>972000</b>      | <b>6.46</b>    | 968090             | 6.43            |
| zaxxon              | <b>725853.9</b>   | <b>867.51</b>  | 18295     | 21.83        | N/A     | N/A             | 109140             | 130.41         | 216020             | 258.15          |
| MEAN HWRNS(%)       |                   | 152.1          |           | 37.9         |         | 4.67            |                    | 117.99         |                    | <b>154.27</b>   |
| Learning Efficiency |                   | 7.61E-11       |           | 1.89E-09     |         | <b>4.67E-08</b> |                    | 5.90E-09       |                    | 7.71E-09        |
| MEDIAN HWRNS(%)     |                   | 49.8           |           | 4.22         |         | 0.13            |                    | <b>35.78</b>   |                    | 50.63           |
| Learning Efficiency |                   | 2.49E-11       |           | 2.11E-10     |         | 1.25E-09        |                    | 1.79E-09       |                    | <b>2.53E-09</b> |
| HWRB                |                   | 19             |           | 3            |         | 0               |                    | 17             |                    | <b>22</b>       |

Table 9: Score table of SOTA model-based algorithms on HWRNS.

| Games               | Muesli        | HWRNS(%)     | Go-Explore    | HWRNS(%)      | GDI-I <sup>3</sup> | HWRNS(%)       | GDI-H <sup>3</sup> | HWRNS(%)        |
|---------------------|---------------|--------------|---------------|---------------|--------------------|----------------|--------------------|-----------------|
| Scale               |               | 200M         | 10B           |               | 200M               |                | 200M               |                 |
| alien               | 139409        | 55.30        | <b>959312</b> | <b>381.06</b> | 43384              | 17.15          | 48735              | 19.27           |
| amidar              | <b>21653</b>  | <b>20.78</b> | 19083         | 18.32         | 1442               | 1.38           | 1065               | 1.02            |
| assault             | 36963         | 436.11       | 30773         | 362.64        | 63876              | 755.57         | <b>97155</b>       | <b>1150.59</b>  |
| asterix             | 316210        | 31.61        | 999500        | 99.95         | 759910             | 75.99          | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 484609        | 4.61         | 112952        | 1.07          | 751970             | 7.15           | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1363427       | 12.75        | 286460        | 2.58          | 3803000            | 35.78          | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | 1213          | 1.46         | <b>3668</b>   | <b>4.45</b>   | 1401               | 1.69           | 1380               | 1.66            |
| battle zone         | 414107        | 51.68        | <b>998800</b> | <b>124.70</b> | 478830             | 59.77          | 824360             | 102.92          |
| beam rider          | 288870        | 28.86        | 371723        | 37.15         | 162100             | 16.18          | <b>422390</b>      | <b>42.22</b>    |
| berzerk             | 44478         | 4.19         | <b>131417</b> | <b>12.41</b>  | 7607               | 0.71           | 14649              | 1.37            |
| bowling             | 191           | 60.64        | <b>247</b>    | <b>80.86</b>  | 202                | 64.57          | 205.2              | 65.76           |
| boxing              | 99            | 99.00        | 91            | 90.99         | <b>100</b>         | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>   |
| breakout            | 791           | 91.53        | 774           | 89.56         | <b>864</b>         | <b>100.00</b>  | <b>864</b>         | <b>100.00</b>   |
| centipede           | <b>869751</b> | <b>66.76</b> | 613815        | 47.07         | 155830             | 11.83          | 195630             | 14.89           |
| chopper command     | 101289        | 10.06        | 996220        | 99.62         | <b>999999</b>      | <b>100.00</b>  | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 175322        | 78.68        | 235600        | 107.51        | 201000             | 90.96          | <b>241170</b>      | <b>110.17</b>   |
| defender            | 629482        | 10.43        | N/A           | N/A           | 893110             | 14.82          | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 129544        | 8.31         | 239895        | 15.41         | 675530             | 43.40          | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | -3            | 39.39        | <b>24</b>     | <b>107.58</b> | <b>24</b>          | <b>107.58</b>  | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2362          | 24.86        | 1031          | 10.85         | <b>14330</b>       | <b>150.84</b>  | 14300              | 150.53          |
| fishing derby       | 51            | 87.71        | <b>67</b>     | <b>97.54</b>  | 59                 | 92.89          | 65                 | 96.31           |
| freeway             | 33            | 86.84        | <b>34</b>     | <b>89.47</b>  | <b>34</b>          | <b>89.47</b>   | <b>34</b>          | <b>89.47</b>    |
| frostbite           | 301694        | 66.33        | <b>999990</b> | <b>219.88</b> | 10485              | 2.29           | 11330              | 2.48            |
| gopher              | 104441        | 29.37        | 134244        | 37.77         | <b>488830</b>      | <b>137.71</b>  | 473560             | 133.41          |
| gravitar            | 11660         | 7.06         | <b>13385</b>  | <b>8.12</b>   | 5905               | 3.52           | 5915               | 3.53            |
| hero                | 37161         | 3.62         | 37783         | 3.68          | <b>38330</b>       | <b>3.73</b>    | 38225              | 3.72            |
| ice hockey          | 25            | 76.69        | 33            | 93.64         | 45                 | 118.94         | <b>47.11</b>       | <b>123.54</b>   |
| jamesbond           | 19319         | 42.38        | 200810        | 441.07        | 594500             | 1305.93        | <b>620780</b>      | <b>1363.66</b>  |
| kangaroo            | 14096         | 0.99         | <b>24300</b>  | <b>1.70</b>   | 14500              | 1.01           | 14636              | 1.02            |
| krull               | 34221         | 31.83        | 63149         | 60.05         | 97575              | 93.63          | <b>594540</b>      | <b>578.47</b>   |
| kung fu master      | 134689        | 13.45        | 24320         | 2.41          | 140440             | 14.02          | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 2359          | 0.19         | <b>24758</b>  | <b>2.03</b>   | 3000               | 0.25           | 2500               | 0.21            |
| ms pacman           | 65278         | 22.42        | <b>456123</b> | <b>157.30</b> | 11536              | 3.87           | 11573              | 3.89            |
| name this game      | 105043        | 448.15       | <b>212824</b> | <b>918.24</b> | 34434              | 140.19         | 36296              | 148.31          |
| phoenix             | 805305        | 20.05        | 19200         | 0.46          | 894460             | 22.27          | <b>959580</b>      | <b>23.89</b>    |
| pitfall             | 0             | 0.20         | <b>7875</b>   | <b>7.09</b>   | 0                  | 0.2            | -4.3               | 0.20            |
| pong                | 20            | 97.60        | <b>21</b>     | <b>100.00</b> | <b>21</b>          | <b>100</b>     | <b>21</b>          | <b>100.00</b>   |
| private eye         | 10323         | 10.12        | <b>69976</b>  | <b>68.73</b>  | 15100              | 14.81          | 15100              | 14.81           |
| qbert               | 157353        | 6.55         | <b>999975</b> | <b>41.66</b>  | 27800              | 1.15           | 28657              | 1.19            |
| riverraid           | <b>47323</b>  | <b>4.60</b>  | 35588         | 3.43          | 28075              | 2.68           | 28349              | 2.70            |
| road runner         | 327025        | 16.05        | 999900        | 49.06         | 878600             | 43.11          | <b>999999</b>      | <b>49.06</b>    |
| robotank            | 59            | 76.96        | <b>143</b>    | <b>190.79</b> | 108                | 143.63         | 113.4              | 150.68          |
| sequest             | 815970        | 81.60        | 539456        | 53.94         | 943910             | 94.39          | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -18407        | -9.47        | <b>-4185</b>  | <b>93.40</b>  | -6774              | 74.67          | -6025              | 86.77           |
| solaris             | 3031          | 1.63         | <b>20306</b>  | <b>17.31</b>  | 11074              | 8.93           | 9105               | 7.14            |
| space invaders      | 59602         | 9.57         | 93147         | 14.97         | 140460             | 22.58          | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 214383        | 278.51       | 609580        | 793.52        | 465750             | 606.09         | <b>677590</b>      | <b>882.15</b>   |
| surround            | <b>9</b>      | <b>96.94</b> | N/A           | N/A           | -8                 | 11.22          | 2.606              | 64.32           |
| tennis              | 12            | 79.91        | <b>24</b>     | <b>106.7</b>  | <b>24</b>          | <b>106.70</b>  | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 359105        | 575.94       | 183620        | 291.67        | 216770             | 345.37         | <b>450810</b>      | <b>724.49</b>   |
| tutankham           | 252           | 4.48         | <b>528</b>    | <b>9.62</b>   | 424                | 7.68           | 418.2              | 7.57            |
| up n down           | 649190        | 788.10       | 553718        | 672.10        | <b>986440</b>      | <b>1197.85</b> | 966590             | 1173.73         |
| venture             | 2104          | 5.41         | <b>3074</b>   | <b>7.90</b>   | 2035               | 5.23           | 2000               | 5.14            |
| video pinball       | 685436        | 0.77         | <b>999999</b> | <b>1.12</b>   | 925830             | 1.04           | 978190             | 1.10            |
| wizard of wor       | 93291         | 23.49        | <b>199900</b> | <b>50.50</b>  | 64293              | 16.14          | 63735              | 16.00           |
| yars revenge        | 557818        | 3.70         | <b>999998</b> | <b>6.65</b>   | 972000             | 6.46           | 968090             | 6.43            |
| zaxxon              | 65325         | 78.04        | 18340         | 21.88         | 109140             | 130.41         | <b>216020</b>      | <b>258.15</b>   |
| MEAN HWRNS(%)       |               | 75.52        |               | 116.89        |                    | 117.99         |                    | <b>154.27</b>   |
| Learning Efficiency |               | 3.78E-09     |               | 1.17E-10      |                    | 5.90E-09       |                    | <b>7.71E-09</b> |
| MEDIAN HWRNS(%)     |               | 24.68        |               | 50.5          |                    | 35.78          |                    | <b>50.63</b>    |
| Learning Efficiency |               | 1.24E-09     |               | 5.05E-11      |                    | 1.79E-09       |                    | <b>2.53E-09</b> |
| HWRB                |               | 5            |               | 15            |                    | 17             |                    | <b>22</b>       |

Table 10: Score table of other SOTA algorithms on HWRNS.

## Atari Games Table of Scores Based on SABER

In this part, we detail the raw score of several representative SOTA algorithms including the SOTA 200M model-free algorithms, SOTA 10B+ model-free algorithms, SOTA model-based algorithms and other SOTA algorithms.<sup>3</sup> Additionally, we calculate the capped human world records normalized world score (CHWRNS) or called SABER (Toromanoff, Wirbel, and Moutarde 2019) of each game with each algorithms. First of all, we demonstrate the sources of the scores that we used. Random scores are from (Badia et al. 2020a). Human world records (HWR) are from (Hafner et al. 2020; Toromanoff, Wirbel, and Moutarde 2019). Rainbow’s scores are from (Hessel et al. 2017). IMPALA’s scores are from (Espeholt et al. 2018). LASER’s scores are from (Schmitt, Hessel, and Simonyan 2020), no sweep at 200M. As there are many versions of R2D2 and NGU, we use original papers’. R2D2’s scores are from (Kapturowski et al. 2018). NGU’s scores are from (Badia et al. 2020b). Agent57’s scores are from (Badia et al. 2020a). MuZero’s scores are from (Schrittwieser et al. 2020). DreamerV2’s scores are from (Hafner et al. 2020). SimPLe’s scores are from (Kaiser et al. 2019). Go-Explore’s scores are from (Ecoffet et al. 2019). Muesli’s scores are from (Hessel et al. 2021). In the following we detail the raw scores and SABER of each algorithms on 57 Atari games.

---

<sup>3</sup>200M and 10B+ represent the training scale.

| Games               | RND     | HWR             | RAINBOW    | SABER(%)      | IMPALA    | SABER(%) | LASER      | SABER(%)      | GDI-I <sup>3</sup> | SABER(%)      | GDI-H <sup>3</sup> | SABER(%)        |
|---------------------|---------|-----------------|------------|---------------|-----------|----------|------------|---------------|--------------------|---------------|--------------------|-----------------|
| Scale               |         |                 | 200M       |               | 200M      |          | 200M       |               | 200M               |               | 200M               |                 |
| alien               | 227.8   | <b>251916</b>   | 9491.7     | 3.68          | 15962.1   | 6.25     | 976.51     | 14.04         | 43384              | 17.15         | 48735              | 19.27           |
| amidar              | 5.8     | <b>104159</b>   | 5131.2     | 4.92          | 1554.79   | 1.49     | 1829.2     | 1.75          | 1442               | 1.38          | 1065               | 1.02            |
| assault             | 222.4   | 8647            | 14198.5    | 165.90        | 19148.47  | 200.00   | 21560.4    | 200.00        | 63876              | 200.00        | <b>97155</b>       | <b>200.00</b>   |
| asterix             | 210     | <b>1000000</b>  | 428200     | 42.81         | 300732    | 30.06    | 240090     | 23.99         | 759910             | 75.99         | 999999             | 100.00          |
| asteroids           | 719     | <b>10506650</b> | 2712.8     | 0.02          | 108590.05 | 1.03     | 213025     | 2.02          | 751970             | 7.15          | 760005             | 7.23            |
| atlantis            | 12850   | <b>10604840</b> | 826660     | 7.68          | 849967.5  | 7.90     | 841200     | 7.82          | 3803000            | 35.78         | 3837300            | 36.11           |
| bank heist          | 14.2    | <b>82058</b>    | 1358       | 1.64          | 1223.15   | 1.47     | 569.4      | 0.68          | 1401               | 1.69          | 1380               | 1.66            |
| battle zone         | 236     | 801000          | 62010      | 7.71          | 20885     | 2.58     | 64953.3    | 8.08          | 478830             | 59.77         | <b>824360</b>      | <b>102.92</b>   |
| beam rider          | 363.9   | <b>999999</b>   | 16850.2    | 1.65          | 32463.47  | 3.21     | 90881.6    | 9.06          | 162100             | 16.18         | 422390             | 42.22           |
| berzerk             | 123.7   | <b>1057940</b>  | 2545.6     | 0.23          | 1852.7    | 0.16     | 25579.5    | 2.41          | 7607               | 0.71          | 14649              | 1.37            |
| bowling             | 23.1    | <b>300</b>      | 30         | 2.49          | 59.92     | 13.30    | 48.3       | 9.10          | 201.9              | 64.57         | 205.2              | 65.76           |
| boxing              | 0.1     | <b>100</b>      | 99.6       | 99.60         | 99.96     | 99.96    | <b>100</b> | <b>100.00</b> | <b>100</b>         | <b>100.00</b> | <b>100</b>         | <b>100.00</b>   |
| breakout            | 1.7     | <b>864</b>      | 417.5      | 48.22         | 787.34    | 91.11    | 747.9      | 86.54         | <b>864</b>         | <b>100.00</b> | <b>864</b>         | <b>100.00</b>   |
| centipede           | 2090.9  | <b>1301709</b>  | 8167.3     | 0.47          | 11049.75  | 0.69     | 292792     | 22.37         | 155830             | 11.83         | 195630             | 14.89           |
| chopper command     | 811     | <b>999999</b>   | 16654      | 1.59          | 28255     | 2.75     | 761699     | 76.15         | <b>999999</b>      | <b>100.00</b> | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 10780.5 | 219900          | 168788.5   | 75.56         | 136950    | 60.33    | 167820     | 75.10         | 201000             | 90.96         | <b>241170</b>      | <b>110.17</b>   |
| defender            | 2874.5  | <b>6010500</b>  | 55105      | 0.87          | 185203    | 3.03     | 336953     | 5.56          | 893110             | 14.82         | 970540             | 16.11           |
| demon attack        | 152.1   | <b>1556345</b>  | 111185     | 7.13          | 132826.98 | 8.53     | 133530     | 8.57          | 675530             | 43.10         | 787985             | 50.63           |
| double dunk         | -18.6   | 21              | -0.3       | 46.21         | -0.33     | 46.14    | 14         | 82.32         | <b>24</b>          | <b>107.58</b> | <b>24</b>          | <b>107.58</b>   |
| enduro              | 0       | 9500            | 2125.9     | 22.38         | 0         | 0.00     | 0          | 0.00          | <b>14330</b>       | <b>150.84</b> | 14300              | 150.53          |
| fishing derby       | -91.7   | <b>71</b>       | 31.3       | 75.60         | 44.85     | 83.93    | 45.2       | 84.14         | 59                 | 95.08         | 65                 | 96.31           |
| freeway             | 0       | <b>38</b>       | 34         | 89.47         | 0         | 0.00     | 0          | 0.00          | 34                 | 89.47         | 34                 | 89.47           |
| frostbite           | 65.2    | <b>454830</b>   | 9590.5     | 2.09          | 317.75    | 0.06     | 5083.5     | 1.10          | 10485              | 2.29          | 11330              | 2.48            |
| gopher              | 257.6   | 355040          | 70354.6    | 19.76         | 66782.3   | 18.75    | 114820.7   | 32.29         | <b>488830</b>      | <b>137.71</b> | 473560             | 133.41          |
| gravitar            | 173     | <b>162850</b>   | 1419.3     | 0.77          | 359.5     | 0.11     | 1106.2     | 0.57          | 5905               | 3.52          | 5915               | 3.53            |
| hero                | 1027    | <b>1000000</b>  | 55887.4    | 5.49          | 33730.55  | 3.27     | 31628.7    | 3.06          | 38330              | 3.73          | 38225              | 3.72            |
| ice hockey          | -11.2   | 36              | 1.1        | 26.06         | 3.48      | 31.10    | 17.4       | 60.59         | 44.92              | 118.94        | <b>47.11</b>       | <b>123.54</b>   |
| jamesbond           | 29      | 45550           | 19809      | 43.45         | 601.5     | 1.26     | 37999.8    | 83.41         | 594500             | 200.00        | <b>620780</b>      | <b>200.00</b>   |
| kangaroo            | 52      | <b>1424600</b>  | 14637.5    | 1.02          | 1632      | 0.11     | 14308      | 1.00          | 14500              | 1.01          | 14636              | 1.02            |
| krull               | 1598    | 104100          | 8741.5     | 6.97          | 8147.4    | 6.39     | 9387.5     | 7.60          | 97575              | 93.63         | <b>594540</b>      | <b>200.00</b>   |
| kung fu master      | 258.5   | 1000000         | 52181      | 5.19          | 43375.5   | 4.31     | 607443     | 60.73         | 140440             | 14.02         | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 0       | <b>1219200</b>  | 384        | 0.03          | 0         | 0.00     | 0.3        | 0.00          | 3000               | 0.25          | 2500               | 0.21            |
| ms pacman           | 307.3   | <b>290090</b>   | 5380.4     | 1.75          | 7342.32   | 2.43     | 6565.5     | 2.16          | 11536              | 3.87          | 11573              | 3.89            |
| name this game      | 2292.3  | 25220           | 13136      | 47.30         | 21537.2   | 83.94    | 26219.5    | 104.36        | 34434              | 140.19        | <b>36296</b>       | <b>148.31</b>   |
| phoenix             | 761.5   | <b>4014440</b>  | 108529     | 2.69          | 210996.45 | 5.24     | 519304     | 12.92         | 894460             | 22.27         | 959580             | 23.89           |
| pitfall             | -229.4  | <b>114000</b>   | <b>0</b>   | <b>0.20</b>   | -1.66     | 0.20     | -0.6       | 0.20          | <b>0</b>           | 0.20          | -4.3               | 0.20            |
| pong                | -20.7   | <b>21</b>       | 20.9       | 99.76         | 20.98     | 99.95    | <b>21</b>  | <b>100.00</b> | <b>21</b>          | <b>100.00</b> | <b>21</b>          | <b>100.00</b>   |
| private eye         | 24.9    | <b>101800</b>   | 4234       | 4.14          | 98.5      | 0.07     | 96.3       | 0.07          | 15100              | 14.81         | 15100              | 14.81           |
| qbert               | 163.9   | <b>2400000</b>  | 33817.5    | 1.40          | 351200.12 | 14.63    | 21449.6    | 0.89          | 27800              | 1.03          | 28657              | 1.19            |
| riverraid           | 1338.5  | <b>1000000</b>  | 22920.8    | 2.16          | 29608.05  | 2.83     | 40362.7    | 3.91          | 28075              | 2.68          | 28349              | 2.70            |
| road runner         | 11.5    | <b>2038100</b>  | 62041      | 3.04          | 57121     | 2.80     | 45289      | 2.22          | 878600             | 43.11         | 999999             | 49.06           |
| robotank            | 2.2     | 76              | 61.4       | 80.22         | 12.96     | 14.58    | 62.1       | 81.17         | 108.2              | 143.63        | <b>113.4</b>       | <b>150.68</b>   |
| seaquest            | 68.4    | 999999          | 15898.9    | 1.58          | 1753.2    | 0.17     | 2890.3     | 0.28          | 943910             | 94.39         | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -17098  | <b>-3272</b>    | -12957.8   | 29.95         | -10180.38 | 50.03    | -29968.4   | -93.09        | -6774              | 74.67         | -6025              | 86.77           |
| solaris             | 1236.3  | <b>111420</b>   | 3560.3     | 2.11          | 2365      | 1.02     | 2273.5     | 0.94          | 11074              | 8.93          | 9105               | 7.14            |
| space invaders      | 148     | <b>621535</b>   | 18789      | 3.00          | 43595.78  | 6.99     | 51037.4    | 8.19          | 140460             | 22.58         | 154380             | 24.82           |
| star gunner         | 664     | 77400           | 127029     | 164.67        | 200625    | 200.00   | 321528     | 418.14        | 465750             | 200.00        | <b>677590</b>      | <b>200.00</b>   |
| surround            | -10     | 9.6             | <b>9.7</b> | <b>100.51</b> | 7.56      | 89.59    | 8.4        | 93.88         | -7.8               | 11.22         | 2.606              | 64.32           |
| tennis              | -23.8   | 21              | 0          | 53.13         | 0.55      | 54.35    | 12.2       | 80.36         | <b>24</b>          | <b>106.70</b> | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 3568    | 65300           | 12926      | 15.16         | 48481.5   | 72.76    | 105316     | 164.82        | 216770             | 200.00        | <b>450810</b>      | <b>200.00</b>   |
| tutankham           | 11.4    | <b>5384</b>     | 241        | 4.27          | 292.11    | 5.22     | 278.9      | 4.98          | 423.9              | 7.68          | 418.2              | 7.57            |
| up n down           | 533.4   | 82840           | 125755     | 152.14        | 332546.75 | 200.00   | 345727     | 200.00        | <b>986440</b>      | <b>200.00</b> | 966590             | 200.00          |
| venture             | 0       | <b>38900</b>    | 5.5        | 0.01          | 0         | 0.00     | 0          | 0.00          | 2000               | 5.14          | 2000               | 5.14            |
| video pinball       | 0       | <b>89218328</b> | 533936.5   | 0.60          | 572898.27 | 0.64     | 511835     | 0.57          | 925830             | 1.04          | 978190             | 1.10            |
| wizard of wor       | 563.5   | <b>395300</b>   | 17862.5    | 4.38          | 9157.5    | 2.18     | 29059.3    | 7.22          | 64439              | 16.18         | 63735              | 16.00           |
| yars revenge        | 3092.9  | <b>15000105</b> | 102557     | 0.66          | 84231.14  | 0.54     | 166292.3   | 1.09          | 972000             | 6.46          | 968090             | 6.43            |
| zaxxon              | 32.5    | 83700           | 22209.5    | 26.51         | 32935.5   | 39.33    | 41118      | 49.11         | 109140             | 130.41        | <b>216020</b>      | <b>200.00</b>   |
| MEAN SABER(%)       | 0.00    | <b>100.00</b>   |            | 28.39         |           | 29.45    |            | 36.78         |                    | 61.66         |                    | 71.26           |
| Learning Efficiency | 0.00    | N/A             |            | 1.42E-09      |           | 1.47E-09 |            | 1.84E-09      |                    | 3.08E-09      |                    | <b>3.56E-09</b> |
| MEDIAN SABER(%)     | 0.00    | <b>100.00</b>   |            | 4.92          |           | 4.31     |            | 8.08          |                    | 35.78         |                    | 50.63           |
| Learning Efficiency | 0.00    | N/A             |            | 2.46E-10      |           | 2.16E-10 |            | 4.04E-10      |                    | 2.27E-09      |                    | <b>2.53E-09</b> |
| HWRB                | 0       | <b>57</b>       |            | 4             |           | 3        |            | 7             |                    | 17            |                    | 22              |

Table 11: Score table of SOTA 200M model-free algorithms on SABER.

| Games               | R2D2           | SABER(%)      | NGU            | SABER(%) | AGENT57          | SABER(%)      | GDI-I <sup>3</sup> | SABER(%)      | GDI-H <sup>3</sup> | SABER(%)        |
|---------------------|----------------|---------------|----------------|----------|------------------|---------------|--------------------|---------------|--------------------|-----------------|
| Scale               | 10B            |               | 35B            |          | 100B             |               | 200M               |               | 200M               |                 |
| alien               | 109038.4       | 43.23         | 248100         | 98.48    | <b>297638.17</b> | <b>118.17</b> | 43384              | 17.15         | 48735              | 19.27           |
| amidar              | 27751.24       | 26.64         | 17800          | 17.08    | <b>29660.08</b>  | <b>28.47</b>  | 1442               | 1.38          | 1065               | 1.02            |
| assault             | 90526.44       | 200.00        | 34800          | 200.00   | 67212.67         | 200.00        | 63876              | 200.00        | <b>97155</b>       | <b>200.00</b>   |
| asterix             | 999080         | 99.91         | 950700         | 95.07    | 991384.42        | 99.14         | 759910             | 75.99         | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 265861.2       | 2.52          | 230500         | 2.19     | 150854.61        | 1.43          | 751970             | 7.15          | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1576068        | 14.76         | 1653600        | 15.49    | 1528841.76       | 14.31         | 3803000            | 35.78         | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | <b>46285.6</b> | <b>56.40</b>  | 17400          | 21.19    | 23071.5          | 28.10         | 1401               | 1.69          | 1380               | 1.66            |
| battle zone         | 513360         | 64.08         | 691700         | 86.35    | <b>934134.88</b> | <b>116.63</b> | 478830             | 59.77         | 824360             | 102.92          |
| beam rider          | 128236.08      | 12.79         | 63600          | 6.33     | 300509.8         | 30.03         | 162100             | 16.18         | <b>422390</b>      | <b>42.22</b>    |
| berzerk             | 34134.8        | 3.22          | 36200          | 3.41     | <b>61507.83</b>  | <b>5.80</b>   | 7607               | 0.71          | 14649              | 1.37            |
| bowling             | 196.36         | 62.57         | 211.9          | 68.18    | <b>251.18</b>    | <b>82.37</b>  | 201.9              | 64.57         | 205.2              | 65.76           |
| boxing              | 99.16          | 99.16         | 99.7           | 99.70    | <b>100</b>       | <b>100.00</b> | <b>100</b>         | <b>100.00</b> | <b>100</b>         | <b>100.00</b>   |
| breakout            | 795.36         | 92.04         | 559.2          | 64.65    | 790.4            | 91.46         | <b>864</b>         | <b>100.00</b> | <b>864</b>         | <b>100</b>      |
| centipede           | 532921.84      | 40.85         | <b>577800</b>  | 44.30    | 412847.86        | 31.61         | 155830             | 11.83         | 195630             | 14.89           |
| chopper command     | 960648         | 96.06         | 999900         | 99.99    | 999900           | 99.99         | <b>999999</b>      | <b>100.00</b> | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 312768         | 144.41        | 313400         | 144.71   | <b>565909.85</b> | <b>200.00</b> | 201000             | 90.96         | 241170             | 110.17          |
| defender            | 562106         | 9.31          | 664100         | 11.01    | 677642.78        | 11.23         | 893110             | 14.82         | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 143664.6       | 9.22          | 143500         | 9.21     | 143161.44        | 9.19          | 675530             | 43.10         | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | 23.12          | 105.35        | -14.1          | 11.36    | 23.93            | 107.40        | <b>24</b>          | <b>107.58</b> | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2376.68        | 25.02         | 2000           | 21.05    | 2367.71          | 24.92         | <b>14330</b>       | <b>150.84</b> | 14300              | 150.53          |
| fishing derby       | 81.96          | 106.74        | 32             | 76.03    | <b>86.97</b>     | <b>109.82</b> | 59                 | 95.08         | 65                 | 96.31           |
| freeway             | <b>34</b>      | <b>89.47</b>  | 28.5           | 75.00    | 32.59            | 85.76         | <b>34</b>          | <b>89.47</b>  | <b>34</b>          | <b>89.47</b>    |
| frostbite           | 11238.4        | 2.46          | 206400         | 45.37    | <b>541280.88</b> | <b>119.01</b> | 10485              | 2.29          | 11330              | 2.48            |
| gopher              | 122196         | 34.37         | 113400         | 31.89    | 117777.08        | 33.12         | <b>488830</b>      | <b>137.71</b> | 473560             | 133.41          |
| gravitar            | 6750           | 4.04          | 14200          | 8.62     | <b>19213.96</b>  | <b>11.70</b>  | 5905               | 3.52          | 5915               | 3.53            |
| hero                | 37030.4        | 3.60          | 69400          | 6.84     | <b>114736.26</b> | <b>11.38</b>  | 38330              | 3.73          | 38225              | 3.72            |
| ice hockey          | <b>71.56</b>   | <b>175.34</b> | -4.1           | 15.04    | 63.64            | 158.56        | 44.92              | 118.94        | <b>47.11</b>       | <b>123.54</b>   |
| jamesbond           | 23266          | 51.05         | 26600          | 58.37    | 135784.96        | 200.00        | 594500             | 200.00        | <b>620780</b>      | <b>200.00</b>   |
| kangaroo            | 14112          | 0.99          | <b>35100</b>   | 2.46     | 24034.16         | 1.68          | 14500              | 1.01          | 14636              | 1.02            |
| krull               | 145284.8       | 140.18        | 127400         | 122.73   | 251997.31        | 200.00        | 97575              | 93.63         | <b>594540</b>      | <b>200.00</b>   |
| kung fu master      | 200176         | 20.00         | 212100         | 21.19    | 206845.82        | 20.66         | 140440             | 14.02         | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 2504           | 0.21          | <b>10400</b>   | 0.85     | 9352.01          | 0.77          | 3000               | 0.25          | 2500               | 0.21            |
| ms pacman           | 29928.2        | 10.22         | 40800          | 13.97    | <b>63994.44</b>  | <b>21.98</b>  | 11536              | 3.87          | 11573              | 3.89            |
| name this game      | 45214.8        | 187.21        | 23900          | 94.24    | <b>54386.77</b>  | <b>200.00</b> | 34434              | 140.19        | 36296              | 148.31          |
| phoenix             | 811621.6       | 20.20         | 959100         | 23.88    | 908264.15        | 22.61         | 894460             | 22.27         | <b>959580</b>      | <b>23.89</b>    |
| pitfall             | <b>0</b>       | <b>0.20</b>   | 7800           | 7.03     | <b>18756.01</b>  | <b>16.62</b>  | 0                  | 0.20          | -4.3               | 0.20            |
| pong                | <b>21</b>      | <b>100.00</b> | 19.6           | 96.64    | 20.67            | 99.21         | <b>21</b>          | <b>100.00</b> | <b>21</b>          | <b>100.00</b>   |
| private eye         | 300            | 0.27          | <b>100000</b>  | 98.23    | 79716.46         | 78.30         | 15100              | 14.81         | 15100              | 14.81           |
| qbert               | 161000         | 6.70          | 451900         | 18.82    | <b>580328.14</b> | <b>24.18</b>  | 27800              | 1.03          | 28657              | 1.19            |
| riverraid           | 34076.4        | 3.28          | 36700          | 3.54     | <b>63318.67</b>  | <b>6.21</b>   | 28075              | 2.68          | 28349              | 2.70            |
| road runner         | 498660         | 24.47         | 128600         | 6.31     | 243025.8         | 11.92         | <b>878600</b>      | <b>43.11</b>  | 999999             | 49.06           |
| robotank            | <b>132.4</b>   | <b>176.42</b> | 9.1            | 9.35     | 127.32           | 169.54        | 108                | 143.63        | 113.4              | 150.68          |
| seaquest            | 999991.84      | 100.00        | <b>1000000</b> | 100.00   | 999997.63        | 100.00        | 943910             | 94.39         | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -29970.32      | -93.10        | -22977.9       | -42.53   | <b>-4202.6</b>   | <b>93.27</b>  | -6774              | 74.67         | -6025              | 86.77           |
| solaris             | 4198.4         | 2.69          | 4700           | 3.14     | <b>44199.93</b>  | <b>38.99</b>  | 11074              | 8.93          | 9105               | 7.14            |
| space invaders      | 55889          | 8.97          | 43400          | 6.96     | 48680.86         | 7.81          | 140460             | 22.58         | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 521728         | 200.00        | 414600         | 200.00   | <b>839573.53</b> | <b>200.00</b> | 465750             | 200.00        | 677590             | 200.00          |
| surround            | <b>9.96</b>    | <b>101.84</b> | -9.6           | 2.04     | 9.5              | 99.49         | -7.8               | 11.22         | 2.606              | 64.32           |
| tennis              | <b>24</b>      | <b>106.70</b> | 10.2           | 75.89    | 23.84            | 106.34        | <b>24</b>          | <b>106.70</b> | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 348932         | 200.00        | 344700         | 200.00   | 405425.31        | 200.00        | 216770             | 200.00        | <b>450810</b>      | <b>200.00</b>   |
| tutankham           | 393.64         | 7.11          | 191.1          | 3.34     | <b>2354.91</b>   | <b>43.62</b>  | 423.9              | 7.68          | 418.2              | 7.57            |
| up n down           | 542918.8       | 200.00        | 620100         | 200.00   | 623805.73        | 200.00        | <b>986440</b>      | <b>200.00</b> | 966590             | 200.00          |
| venture             | 1992           | 5.12          | 1700           | 4.37     | <b>2623.71</b>   | <b>6.74</b>   | 2000               | 5.14          | 2000               | 5.14            |
| video pinball       | 483569.72      | 0.54          | 965300         | 1.08     | <b>992340.74</b> | <b>1.11</b>   | 925830             | 1.04          | 978190             | 1.10            |
| wizard of wor       | 133264         | 33.62         | 106200         | 26.76    | <b>157306.41</b> | <b>39.71</b>  | 64439              | 16.18         | 63735              | 16.00           |
| yars revenge        | 918854.32      | 6.11          | 986000         | 6.55     | <b>998532.37</b> | <b>6.64</b>   | 972000             | 6.46          | 968090             | 6.43            |
| zaxxon              | 181372         | 200.00        | 111100         | 132.75   | <b>249808.9</b>  | <b>200.00</b> | 109140             | 130.41        | 216020             | 200.00          |
| MEAN SABER(%)       |                | 60.43         |                | 50.47    |                  | <b>76.26</b>  |                    | 61.66         |                    | 71.26           |
| Learning Efficiency |                | 6.04E-11      |                | 1.44E-11 |                  | 7.63E-12      |                    | 5.90E-09      |                    | <b>3.56E-09</b> |
| MEDIAN SABER(%)     |                | 33.62         |                | 21.19    |                  | 43.62         |                    | 35.78         |                    | <b>50.63</b>    |
| Learning Efficiency |                | 3.36E-11      |                | 6.05E-12 |                  | 4.36E-12      |                    | 2.27E-09      |                    | <b>2.53E-09</b> |
| HWRB                |                | 15            |                | 9        |                  | 18            |                    | 17            |                    | <b>22</b>       |

Table 12: Score table of SOTA 10B+ model-free algorithms on SABER.

| Games               | MuZero            | SABER(%)      | DreamerV2 | SABER(%)     | SimPLe  | SABER(%)        | GDI-I <sup>3</sup> | SABER(%)      | GDI-H <sup>3</sup> | SABER(%)        |
|---------------------|-------------------|---------------|-----------|--------------|---------|-----------------|--------------------|---------------|--------------------|-----------------|
| Scale               | 20B               |               | 200M      |              | 1M      |                 | 200M               |               | 200M               |                 |
| alien               | <b>741812.63</b>  | <b>200.00</b> | 3483      | 1.29         | 616.9   | 0.15            | 43384              | 17.15         | 48735              | 19.27           |
| amidar              | <b>28634.39</b>   | <b>27.49</b>  | 2028      | 1.94         | 74.3    | 0.07            | 1442               | 1.38          | 1065               | 1.02            |
| assault             | <b>143972.03</b>  | <b>200.00</b> | 7679      | 88.51        | 527.2   | 3.62            | 63876              | 200.00        | 97155              | 200.00          |
| asterix             | 998425            | 99.84         | 25669     | 2.55         | 1128.3  | 0.09            | 759910             | 75.99         | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 678558.64         | 6.45          | 3064      | 0.02         | 793.6   | 0.00            | 751970             | 7.15          | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1674767.2         | 15.69         | 989207    | 9.22         | 20992.5 | 0.08            | 3803000            | 35.78         | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | 1278.98           | 1.54          | 1043      | 1.25         | 34.2    | 0.02            | <b>1401</b>        | <b>1.69</b>   | 1380               | 1.66            |
| battle zone         | <b>848623</b>     | <b>105.95</b> | 31225     | 3.87         | 4031.2  | 0.47            | 478830             | 59.77         | 824360             | 102.92          |
| beam rider          | <b>454993.53</b>  | <b>45.48</b>  | 12413     | 1.21         | 621.6   | 0.03            | 162100             | 16.18         | 422390             | 42.22           |
| berzerk             | <b>85932.6</b>    | <b>8.11</b>   | 751       | 0.06         | N/A     | N/A             | 7607               | 0.71          | 14649              | 1.37            |
| bowling             | <b>260.13</b>     | <b>85.60</b>  | 48        | 8.99         | 30      | 2.49            | 202                | 64.57         | 205.2              | 65.76           |
| boxing              | <b>100</b>        | <b>100.00</b> | 87        | 86.99        | 7.8     | 7.71            | <b>100</b>         | <b>100.00</b> | <b>100</b>         | <b>100.00</b>   |
| breakout            | <b>864</b>        | <b>100.00</b> | 350       | 40.39        | 16.4    | 1.70            | <b>864</b>         | <b>100.00</b> | <b>864</b>         | 100.00          |
| centipede           | <b>1159049.27</b> | <b>89.02</b>  | 6601      | 0.35         | N/A     | N/A             | 155830             | 11.83         | 195630             | 14.89           |
| chopper command     | 991039.7          | 99.10         | 2833      | 0.20         | 979.4   | 0.02            | <b>999999</b>      | <b>100.00</b> | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | <b>458315.4</b>   | <b>200.00</b> | 141424    | 62.47        | 62583.6 | 24.77           | 201000             | 90.96         | 241170             | 110.17          |
| defender            | 839642.95         | 13.93         | N/A       | N/A          | N/A     | N/A             | 893110             | 14.82         | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 143964.26         | 9.24          | 2775      | 0.17         | 208.1   | 0.00            | 675530             | 43.40         | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | 23.94             | 107.42        | 22        | 102.53       | N/A     | N/A             | <b>24</b>          | <b>107.58</b> | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2382.44           | 25.08         | 2112      | 22.23        | N/A     | N/A             | <b>14330</b>       | <b>150.84</b> | 14300              | 150.53          |
| fishing derby       | <b>91.16</b>      | <b>112.39</b> | 93.24     | 200.00       | -90.7   | 0.61            | 59                 | 92.89         | 65                 | 96.31           |
| freeway             | 33.03             | 86.92         | <b>34</b> | <b>89.47</b> | 16.7    | 43.95           | <b>34</b>          | <b>89.47</b>  | <b>34</b>          | <b>89.47</b>    |
| frostbite           | <b>631378.53</b>  | <b>138.82</b> | 15622     | 3.42         | 236.9   | 0.04            | 10485              | 2.29          | 11330              | 2.48            |
| gopher              | 130345.58         | 36.67         | 53853     | 15.11        | 596.8   | 0.10            | <b>488830</b>      | <b>137.71</b> | 473560             | 133.41          |
| gravitar            | <b>6682.7</b>     | <b>4.00</b>   | 3554      | 2.08         | 173.4   | 0.00            | 5905               | 3.52          | 5915               | 3.53            |
| hero                | <b>49244.11</b>   | <b>4.83</b>   | 30287     | 2.93         | 2656.6  | 0.16            | 38330              | 3.73          | 38225              | 3.72            |
| ice hockey          | <b>67.04</b>      | <b>165.76</b> | 29        | 85.17        | -11.6   | -0.85           | 44.92              | 118.94        | 47.11              | 123.54          |
| jamesbond           | 41063.25          | 90.14         | 9269      | 20.30        | 100.5   | 0.16            | 594500             | 200.00        | <b>620780</b>      | <b>200.00</b>   |
| kangaroo            | <b>16763.6</b>    | <b>1.17</b>   | 11819     | 0.83         | 51.2    | 0.00            | 14500              | 1.01          | 14636              | 1.02            |
| krull               | 269358.27         | 200.00        | 9687      | 7.89         | 2204.8  | 0.59            | 97575              | 93.63         | <b>594540</b>      | <b>200.00</b>   |
| kung fu master      | <b>204824</b>     | <b>20.46</b>  | 66410     | 6.62         | 14862.5 | 1.46            | 140440             | 14.02         | 1666665            | 166.68          |
| montezuma revenge   | 0                 | 0.00          | 1932      | 0.16         | N/A     | N/A             | <b>3000</b>        | <b>0.25</b>   | 2500               | 0.21            |
| ms pacman           | <b>243401.1</b>   | <b>83.89</b>  | 5651      | 1.84         | 1480    | 0.40            | 11536              | 3.87          | 11573              | 3.89            |
| name this game      | <b>157177.85</b>  | <b>200.00</b> | 14472     | 53.12        | 2420.7  | 0.56            | 34434              | 140.19        | 36296              | 148.31          |
| phoenix             | <b>955137.84</b>  | <b>23.78</b>  | 13342     | 0.31         | N/A     | N/A             | 894460             | 22.27         | 959580             | 23.89           |
| pitfall             | <b>0</b>          | <b>0.20</b>   | -1        | 0.20         | N/A     | N/A             | <b>0</b>           | <b>0.20</b>   | -4.3               | 0.20            |
| pong                | <b>21</b>         | <b>100.00</b> | 19        | 95.20        | 12.8    | 80.34           | <b>21</b>          | <b>100.00</b> | <b>21</b>          | <b>100.00</b>   |
| private eye         | <b>15299.98</b>   | <b>15.01</b>  | 158       | 0.13         | 35      | 0.01            | 15100              | 14.81         | 15100              | 14.81           |
| qbert               | <b>72276</b>      | <b>3.00</b>   | 162023    | 6.74         | 1288.8  | 0.05            | 27800              | 1.15          | 28657              | 1.19            |
| riverraid           | <b>323417.18</b>  | <b>32.25</b>  | 16249     | 1.49         | 1957.8  | 0.06            | 28075              | 2.68          | 28349              | 2.70            |
| road runner         | 613411.8          | 30.10         | 88772     | 4.36         | 5640.6  | 0.28            | 878600             | 43.11         | <b>999999</b>      | <b>49.06</b>    |
| robotank            | <b>131.13</b>     | <b>174.70</b> | 65        | 85.09        | N/A     | N/A             | 108                | 143.63        | 113.4              | 150.68          |
| seaquest            | 999976.52         | 100.00        | 45898     | 4.58         | 683.3   | 0.06            | 943910             | 94.39         | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | <b>-29968.36</b>  | <b>-93.09</b> | -8187     | 64.45        | N/A     | N/A             | -6774              | 74.67         | -6025              | 86.77           |
| solaris             | 56.62             | -1.07         | 883       | -0.32        | N/A     | N/A             | <b>11074</b>       | <b>8.93</b>   | 9105               | 7.14            |
| space invaders      | 74335.3           | 11.94         | 2611      | 0.40         | N/A     | N/A             | 140460             | 22.58         | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 549271.7          | 200.00        | 29219     | 37.21        | N/A     | N/A             | 465750             | 200.00        | <b>677590</b>      | <b>200.00</b>   |
| surround            | <b>9.99</b>       | <b>101.99</b> | N/A       | N/A          | N/A     | N/A             | -8                 | 11.22         | 2.606              | 64.32           |
| tennis              | 0                 | 53.13         | 23        | 104.46       | N/A     | N/A             | <b>24</b>          | <b>106.70</b> | <b>24</b>          | <b>106.70</b>   |
| time pilot          | <b>476763.9</b>   | <b>200.00</b> | 32404     | 46.71        | N/A     | N/A             | 216770             | 200.00        | 450810             | 200.00          |
| tutankham           | <b>491.48</b>     | <b>8.94</b>   | 238       | 4.22         | N/A     | N/A             | 424                | 7.68          | 418.2              | 7.57            |
| up n down           | 715545.61         | 200.00        | 648363    | 200.00       | 3350.3  | 3.42            | <b>986440</b>      | <b>200.00</b> | 966590             | 200.00          |
| venture             | 0.4               | 0.00          | 0         | 0.00         | N/A     | N/A             | <b>2000</b>        | <b>5.23</b>   | 2000               | 5.14            |
| video pinball       | <b>981791.88</b>  | <b>1.10</b>   | 22218     | 0.02         | N/A     | N/A             | 925830             | 1.04          | 978190             | 1.10            |
| wizard of wor       | <b>197126</b>     | <b>49.80</b>  | 14531     | 3.54         | N/A     | N/A             | 64439              | 16.14         | 63735              | 16.00           |
| yars revenge        | 553311.46         | 3.67          | 20089     | 0.11         | 5664.3  | 0.02            | <b>972000</b>      | <b>6.46</b>   | 968090             | 6.43            |
| zaxxon              | <b>725853.9</b>   | <b>200.00</b> | 18295     | 21.83        | N/A     | N/A             | 109140             | 130.41        | 216020             | 200.00          |
| MEAN SABER(%)       |                   | <b>71.94</b>  |           | 27.22        |         | 4.67            |                    | 61.66         |                    | 71.26           |
| Learning Efficiency |                   | 3.60E-11      |           | 1.36E-09     |         | <b>4.67E-08</b> |                    | 5.90E-09      |                    | 3.56E-09        |
| MEDIAN SABER(%)     |                   | 49.8          |           | 4.22         |         | 0.13            |                    | 35.78         |                    | <b>50.63</b>    |
| Learning Efficiency |                   | 2.49E-11      |           | 2.11E-10     |         | 1.60E-09        |                    | 2.27E-09      |                    | <b>2.53E-09</b> |
| HWRB                |                   | 19            |           | 3            |         | 0               |                    | 17            |                    | <b>22</b>       |

Table 13: Score table of SOTA model-based algorithms on SABER.

| Games               | Muesli        | SABER(%)     | Go-Explore    | SABER(%)      | GDI-I <sup>3</sup> | SABER(%)       | GDI-H <sup>3</sup> | SABER(%)        |
|---------------------|---------------|--------------|---------------|---------------|--------------------|----------------|--------------------|-----------------|
| Scale               | 200M          |              | 10B           |               | 200M               |                | 200M               |                 |
| alien               | 139409        | 55.30        | <b>959312</b> | <b>200.00</b> | 43384              | 17.15          | 48735              | 19.27           |
| amidar              | <b>21653</b>  | <b>20.78</b> | 19083         | 18.32         | 1442               | 1.38           | 1065               | 1.02            |
| assault             | 36963         | 200.00       | 30773         | 200.00        | 63876              | 200.00         | <b>97155</b>       | <b>200.00</b>   |
| asterix             | 316210        | 31.61        | 999500        | 99.95         | 759910             | 75.99          | <b>999999</b>      | <b>100.00</b>   |
| asteroids           | 484609        | 4.61         | 112952        | 1.07          | 751970             | 7.15           | <b>760005</b>      | <b>7.23</b>     |
| atlantis            | 1363427       | 12.75        | 286460        | 2.58          | 3803000            | 35.78          | <b>3837300</b>     | <b>36.11</b>    |
| bank heist          | 1213          | 1.46         | <b>3668</b>   | <b>4.45</b>   | 1401               | 1.69           | 1380               | 1.66            |
| battle zone         | 414107        | 51.68        | <b>998800</b> | <b>124.70</b> | 478830             | 59.77          | 824360             | 102.92          |
| beam rider          | 288870        | 28.86        | 371723        | 37.15         | 162100             | 16.18          | <b>422390</b>      | <b>42.22</b>    |
| berzerk             | 44478         | 4.19         | <b>131417</b> | <b>12.41</b>  | 7607               | 0.71           | 14649              | 1.37            |
| bowling             | 191           | 60.64        | <b>247</b>    | <b>80.86</b>  | 202                | 64.57          | 205.2              | 65.76           |
| boxing              | 99            | 99.00        | 91            | 90.99         | <b>100</b>         | <b>100.00</b>  | <b>100</b>         | <b>100.00</b>   |
| breakout            | 791           | 91.53        | 774           | 89.56         | <b>864</b>         | <b>100.00</b>  | <b>864</b>         | <b>100.00</b>   |
| centipede           | <b>869751</b> | <b>66.76</b> | 613815        | 47.07         | 155830             | 11.83          | 195630             | 14.89           |
| chopper command     | 101289        | 10.06        | 996220        | 99.62         | <b>999999</b>      | <b>100.00</b>  | <b>999999</b>      | <b>100.00</b>   |
| crazy climber       | 175322        | 78.68        | 235600        | 107.51        | 201000             | 90.96          | <b>241170</b>      | <b>110.17</b>   |
| defender            | 629482        | 10.43        | N/A           | N/A           | 893110             | 14.82          | <b>970540</b>      | <b>16.11</b>    |
| demon attack        | 129544        | 8.31         | 239895        | 15.41         | 675530             | 43.40          | <b>787985</b>      | <b>50.63</b>    |
| double dunk         | -3            | 39.39        | <b>24</b>     | <b>107.58</b> | <b>24</b>          | <b>107.58</b>  | <b>24</b>          | <b>107.58</b>   |
| enduro              | 2362          | 24.86        | 1031          | 10.85         | <b>14330</b>       | <b>150.84</b>  | 14300              | 150.53          |
| fishing derby       | 51            | 87.71        | <b>67</b>     | <b>97.54</b>  | 59                 | 92.89          | 65                 | 96.31           |
| freeway             | 33            | 86.84        | <b>34</b>     | <b>89.47</b>  | <b>34</b>          | <b>89.47</b>   | <b>34</b>          | <b>89.47</b>    |
| frostbite           | 301694        | 66.33        | <b>999990</b> | <b>200.00</b> | 10485              | 2.29           | 11330              | 2.48            |
| gopher              | 104441        | 29.37        | 134244        | 37.77         | <b>488830</b>      | <b>137.71</b>  | 473560             | 133.41          |
| gravitar            | 11660         | 7.06         | <b>13385</b>  | <b>8.12</b>   | 5905               | 3.52           | 5915               | 3.53            |
| hero                | 37161         | 3.62         | 37783         | 3.68          | 38330              | 3.73           | <b>38225</b>       | <b>3.72</b>     |
| ice hockey          | 25            | 76.69        | 33            | 93.64         | 44.92              | 118.94         | <b>47.11</b>       | <b>123.54</b>   |
| jamesbond           | 19319         | 42.38        | 200810        | 200.00        | 594500             | 200.00         | <b>620780</b>      | <b>200.00</b>   |
| kangaroo            | 14096         | 0.99         | <b>24300</b>  | <b>1.70</b>   | 14500              | 1.01           | 14636              | 1.02            |
| krull               | 34221         | 31.83        | 63149         | 60.05         | 97575              | 93.63          | <b>594540</b>      | <b>200.00</b>   |
| kung fu master      | 134689        | 13.45        | 24320         | 2.41          | 140440             | 14.02          | <b>1666665</b>     | <b>166.68</b>   |
| montezuma revenge   | 2359          | 0.19         | <b>24758</b>  | <b>2.03</b>   | 3000               | 0.25           | 2500               | 0.21            |
| ms pacman           | 65278         | 22.42        | <b>456123</b> | <b>157.30</b> | 11536              | 3.87           | 11573              | 3.89            |
| name this game      | 105043        | 200.00       | <b>212824</b> | <b>200.00</b> | 34434              | 140.19         | 36296              | 148.31          |
| phoenix             | 805305        | 20.05        | 19200         | 0.46          | 894460             | 22.27          | <b>959580</b>      | <b>23.89</b>    |
| pitfall             | 0             | 0.20         | <b>7875</b>   | <b>7.09</b>   | 0                  | 0.2            | -4.3               | 0.20            |
| pong                | 20            | 97.60        | <b>21</b>     | <b>100.00</b> | <b>21</b>          | <b>100</b>     | <b>21</b>          | <b>100.00</b>   |
| private eye         | 10323         | 10.12        | <b>69976</b>  | <b>68.73</b>  | 15100              | 14.81          | 15100              | 14.81           |
| qbert               | 157353        | 6.55         | <b>999975</b> | <b>41.66</b>  | 27800              | 1.15           | 28657              | 1.19            |
| riverraid           | <b>47323</b>  | <b>4.60</b>  | 35588         | 3.43          | 28075              | 2.68           | 28349              | 2.70            |
| road runner         | 327025        | 16.05        | 999900        | 49.06         | 878600             | 43.11          | <b>999999</b>      | <b>49.06</b>    |
| robotank            | 59            | 76.96        | <b>143</b>    | <b>190.79</b> | 108                | 143.63         | 113.4              | 150.68          |
| seaquest            | 815970        | 81.60        | 539456        | 53.94         | 943910             | 94.39          | <b>1000000</b>     | <b>100.00</b>   |
| skiing              | -18407        | -9.47        | <b>-4185</b>  | <b>93.40</b>  | -6774              | 74.67          | -6025              | 86.77           |
| solaris             | 3031          | 1.63         | <b>20306</b>  | <b>17.31</b>  | 11074              | 8.93           | 9105               | 7.14            |
| space invaders      | 59602         | 9.57         | 93147         | 14.97         | 140460             | 22.58          | <b>154380</b>      | <b>24.82</b>    |
| star gunner         | 214383        | 200.00       | 609580        | 200.00        | 465750             | 200.00         | <b>677590</b>      | <b>200.00</b>   |
| surround            | <b>9</b>      | <b>96.94</b> | N/A           | N/A           | -8                 | 11.22          | 2.606              | 64.32           |
| tennis              | 12            | 79.91        | <b>24</b>     | <b>106.7</b>  | <b>24</b>          | <b>106.70</b>  | <b>24</b>          | <b>106.70</b>   |
| time pilot          | 359105        | 200.00       | 183620        | 200.00        | 216770             | 200.00         | <b>450810</b>      | <b>200.00</b>   |
| tutankham           | 252           | 4.48         | <b>528</b>    | <b>9.62</b>   | 424                | 7.68           | 418.2              | 7.57            |
| up n down           | 649190        | 200.00       | 553718        | 200.00        | <b>986440</b>      | <b>11.9785</b> | 966590             | 200.00          |
| venture             | 2104          | 5.41         | <b>3074</b>   | <b>7.90</b>   | 2035               | 5.23           | 2000               | 5.14            |
| video pinball       | 685436        | 0.77         | <b>999999</b> | <b>1.12</b>   | 925830             | 1.04           | 978190             | 1.10            |
| wizard of wor       | 93291         | 23.49        | <b>199900</b> | <b>50.50</b>  | 64293              | 16.14          | 63735              | 16.00           |
| yars revenge        | 557818        | 3.70         | <b>999998</b> | <b>6.65</b>   | 972000             | 6.46           | 968090             | 6.43            |
| zaxxon              | 65325         | 78.04        | 18340         | 21.88         | 109140             | 130.41         | <b>216020</b>      | <b>200.00</b>   |
| MEAN SABER(%)       |               | 48.74        |               | <b>71.80</b>  |                    | 61.66          |                    | 71.26           |
| Learning Efficiency |               | 2.43E-09     |               | 7.18E-11      |                    | 3.08E-09       |                    | <b>3.56E-09</b> |
| MEDIAN SABER(%)     |               | 24.86        |               | 50.5          |                    | 35.78          |                    | <b>50.63</b>    |
| Learning Efficiency |               | 1.24E-09     |               | 5.05E-11      |                    | 1.78E-09       |                    | <b>2.53E-09</b> |
| HWRB                |               | 5            |               | 15            |                    | 17             |                    | <b>22</b>       |

Table 14: Score table of other SOTA algorithms on SABER.