

Criterion	Score (0-5 or N/A)	Confidence (0-3)	Comments
Algorithm [1]	3	2	Outlines historical foundation of TD-gammon and shows full algorithm
Program [2]	3	2	7 atari games were used in an emulator - no versions given or info on whether the emulator is publicly available
Compilation [3]	N/A	0	No information was given on how code was compiled
Transformations [4]	N/A	0	No transformation tool was mentioned, but it is fair to assume none was used give the scope of this paper
Binary [5]	N/A	0	Similar to compiler, no information was given about executable
Model [6]	3	3	Model is not included, nor are there instructions on how to install but the model architecture is covered fairly well
Data set [7]	5	2	Link to arcade-learning-environment was provided at arXiv
Run-time environment [8]	0	3	No run-time environment information given
Hardware [9]	1	3	No hardware information given
Run-time state [10]	N/A	0	No run-time environment information given
Execution [11]	0	2	No information given regarding execution of workflow
Metrics [12]	5	2	Training stability is used is measured as a function of both rewards and Q-values
Output [13]	5	2	The output is the average total reward for each of the 7 games played
Experiments [14]	1	2	Only information on hyperparameters is given
Required disk space [15]	N/A	0	
Required workflow preparation time [16]	N/A	0	
Required experimental time [17]	1	2	Authors did not detail how long training/testing took
Publicly available [18]	2	2	Original code is not publicly available, but there are 86 community code implementations
Code licenses [19]	N/A	0	
Workflow framework used [20]	N/A	0	
Archived? [21]	N/A	0	Didn't see any DOI
Score	29		
Total	60		

- [1] Are you presenting a new algorithm? If so, do you provide proofs? Are you using an existing algorithm? Is the relevance of the algorithm discussed?
- [2] Which benchmarks do you use (PARSEC ARM real workloads, NAS, EEMBC, SPLASH, Rodinia, LINPACK, HPCG, MiBench, SPEC, cTuning, etc)? Are they included or should they be downloaded? Which version? Are they public or private? If they are private, is there a public analog to evaluate your artifact? What is the approximate size?
- [3] Do you require a specific compiler? Public/private? Is it included? Which version?
- [4] Do you require a program transformation tool (source-to-source, binary-to-binary, compiler pass, etc)? Public/private? Is it included? Which version?
- [5] Are binaries included? OS-specific? Which version?
- [6] Do you use specific models (ImageNet, AlexNet, MobileNets)? Are they included? If not, how to download and install? What is their approximate size?
- [7] Do you use specific data sets? Are they included? If not, how to download and install? What is their approximate size?
- [8] Is your artifact OS-specific (Linux, Windows, MacOS, Android, etc) ? Which version? Which are the main software dependencies (JIT, libs, run-time adaptation frameworks, etc); Do you need root access?
- [9] Do you need specific hardware (supercomputer, architecture simulator, CPU, GPU, neural network accelerator, FPGA) or specific features (hardware counters to measure power consumption, SUDO access to CPU/GPU frequency, etc)? Are they publicly available?
- [10] Is your artifact sensitive to run-time state (cold/hot cache, network/cache contentions, etc.)
- [11] Any specific conditions should be met during experiments (sole user, process pinning, profiling, adaptation, etc)? How long will it approximately run?
- [12] Which metrics are reported (execution time, inference per second, Top1 accuracy, static and dynamic energy consumption, etc) - particularly important for multi-objective benchmarking, optimization and co-design
- [13] What is your output (console, file, table, graph) and what is your result (exact output, numerical results, measured characteristics, etc)? Is expected result included?
- [14] How to prepare experiments and replicate/reproduce results (OS scripts, manual steps by user, IPython/Jupyter notebook, automated workflows, etc)? Do not forget to mention the maximum allowable variation of empirical results!
- [15] How much disk space required (approximately)?

[16] How much time is needed to prepare workflow (approximately)?

[17] How much time is needed to complete experiments (approximately)?

[18] Will your artifact be publicly available? If yes, we may spend an extra effort to help you with the documentation.

[19] If you workflows and artifacts will be publicly available, please provide information about licenses. This will help the community to reuse your components.

[20] Did authors use any workflow framework which can automate and customize experiments?

[21] Note that the author-created artifacts relevant to this paper will receive the ACM "artifact available" badge *only if* they have been placed on a publicly accessible archival repository such as Zenodo, FigShare or Dryad. A DOI will be then assigned to their artifacts and must be provided here! Personal web pages, Google Drive, GitHub, GitLab and BitBucket are not accepted for this badge. Authors can provide this link at the end of the evaluation.