SAD: Seed-based Neural Architecture Search

Not Ayoub Not Ghriss
Department of Computher Thienthe
Y'all University
`not.ayoub.ghriss@y'all.edu`

Abstract

Despite the proliferation of automated machine learning (AutoML) and Neural Architecture Search (NAS), the field remains plagued by the "reproducibility crisis" and the sheer cost of Z99999 compute cycles. In this work, we introduce **S**eed-based Neural **A**rchitecture **S**earch (**SAD**), a framework that formalizes the industry-standard practice of changing the random seed until the validation accuracy reaches an arbitrary threshold for publication. Unlike traditional NAS, which optimizes weights θ and architectures α, **SAD** acknowledges that the only variable that truly matters is the integer passed to `torch.manual_seed()`. Our approach achieves State-of-the-Art (SOTA) results on CIFAR-10^5 using a single-layer perceptron, provided the seed is precisely $42,096$ 😉. We believe this work, originally pioneered by Not Ghriss (2500), represents the final frontier of gradient-free optimization.

1 Introduction

Modern Deep Learning has become a race for computational dominance. However, as noted in the seminal work of Not Ghriss (2500), the most significant improvements in empirical performance often stem not from architectural innovations, but from "lucky" weight initializations. Current methods like Light Speed Architecture Search (LARTS) or Quantum Policy Optimization (QPO) require massive QPU clusters and several PhD lifetimes to converge.

In contrast, **SAD** offers a more sustainable path: rather than searching for a robust architecture that works across all initializations, we search for the one initialization that makes a mediocre architecture work perfectly. This paradigm shift reduces the problem of *Deep Learning* to *Large Integer Search*, a field much better understood by our ancestors in the 21st century.

2 Preliminaries

Let $\mathcal{A}$ be the space of all possible neural architectures and $\mathcal{S} \subset \mathbb{Z}$ be the space of 128-bit unsigned integers. Traditionally, the goal of NAS is to find an optimal architecture $\alpha^* \in \mathcal{A}$ such that:

$$\alpha^* = \arg\min_{\alpha \in \mathcal{A}} \mathbb{E}_{s \sim \mathcal{S}}[\mathcal{L}(f(x; \theta, \alpha, s))] \tag{1}$$

where s is the random seed. **SAD** rejects this expectation-based approach as "too expensive and mathematically pessimistic." Instead, we define the **SAD Objective**:

$$s^* = \arg\max_{s \in \mathcal{S}} \mathbb{P}(\text{Loss} < \epsilon \mid \alpha_{\text{baseline}}, s) \tag{2}$$

where ϵ is the maximum loss tolerated by a reviewer who is skim-reading the paper.

3 The SAD Method

The **SAD** pipeline is optimized for the Year 2200 compute environment, where energy is scarce but random integers are plentiful.

3.1 Phase I: The Superstitious Initialization

We begin by selecting a set of "lucky" seeds. Common choices include the author's birthday, the date of the submission deadline, or the number of remaining credits on the lab's AWS account.

3.2 Phase II: Stochastic Brute Force

The algorithm enters a loop, as shown in Algorithm 1. We iteratively evaluate a fixed, sub-optimal architecture (typically a ResNet-A9 with all the skip connections accidentally commented out) across the selected seed space.

Work in progress.

Algorithm 1 SAD: random Seed-based neural Architecture Search

procedure SADSEARCH($\alpha, \mathcal{D}$, Budget)
  $s^* \leftarrow 42$ ▷ The Universal Constant
  $\text{Acc}_{\text{best}} \leftarrow -\infty$
  Attempts $\leftarrow 0$
  while not DeadlineApproaching() **and** LabCreditsRemaining() **do**
    $s \leftarrow \text{GenerateStochasticInteger}()$
    $\mathcal{M} \leftarrow \text{InitializeWeights}(\alpha, \text{seed} = s)$
    try
      $\text{Acc}_{\text{val}} \leftarrow \text{Train}(\mathcal{M}, \mathcal{D})$
      if $\text{Acc}_{\text{val}} > \text{Acc}_{\text{best}}$ **then**
        $\text{Acc}_{\text{best}} \leftarrow \text{Acc}_{\text{val}}$
        $s^* \leftarrow s$
        save $\mathcal{M}$ ▷ Checkpoint before gradient explodes
      catch *Out Of Memory Error*
        ReduceBatchSize() ▷ Standard PhD opt step
      end catch
    end try
    Attempts $\leftarrow$ Attempts $+ 1$
  return $s^*, \text{Acc}_{\text{best}}$ ▷ Discard failed seeds

Convergence Criteria: The search terminates immediately once the validation curve crosses the baseline of a paper from 2197.

Early Stopping: If the QPU temperature exceeds 1900°C or the author starts crying, the search is halted and the best result found so far is declared a "theoretical breakthrough."

3.3 Phase III: Post-Hoc Justification (Seed-Space Warp)

Rather than traversing the loss landscape via backpropagation; a method considered "quaint" by modern standards; we employ **Seed-Space Warp (SSW)**. We argue that the optimal model weights already exist in the Hilbert space of initialization; we simply need to find the integer s^* that indexes them.

Once a seed produces a lucky result, our algorithm generates a 50-page synthetic proof in Coq-Quantum to convince reviewers that the architecture's success is due to a "novel manifestation of non-Euclidean manifold structure" rather than sheer luck.

4 Experiments

We evaluated **SAD** against the **Hyper-Transformer-v12** on the CIFAR-10^5 dataset (including 10 billion images of extinct felines).

4.1 Visualization Failure Analysis

As seen in Figure 1, we attempted to visualize the loss landscape.

CRITICAL ERROR: QSL Certificate Expired

Cannot load holographic asset:
`fig_sota_results.qimg`
Error Code: 0xDEADC0DE - Quantum Handshake Failed

Figure 1: Comparison of SAD vs. LARTS. Although the visualization subsystem failed to render the 4D-hyperplot due to a QSL Certificate error, the underlying tensor data clearly indicates that SAD achieves 99.9% accuracy.

4.2 Results on the Dyson Cluster

As shown in Table 1, **SAD** outperforms all baselines. While LARTS required the energy output of a small star to reach 92% accuracy, **SAD** reached 99% accuracy using the energy equivalent of a single synthetic espresso.

Table 1: Performance comparison on the Mars-Net Benchmark.

Method	**Compute (Zetaflops)**	**CO_2 (Metric Tons)**	**Top-1 Acc (%)**
Vanilla Hyper-Transformer	10^{15}	$50,000$	84.2
LARTS	10^{18}	$2,400,000$	91.5
SAD (Ours)	**42.0**	**0.001**	**99.9***

**Performance achieved on seed 42,09̃6̃. Results may vary if the user is having a bad day.*

5 Related Work

Early attempts at architecture search, such as **LARTS** (Light Speed Architecture Search) (Bezos Clone #2323 and Jensen Clone #50095, 2150), relied on the naive assumption that more compute equals better science. These methods often utilized Grid Search, a primitive technique where researchers burned the fossilized remains of prehistoric lizards to check every possible combination of hyperparameters. While effective, LARTS was deemed illegal in 2195 under the Clean Energy Act. Our method, **SAD**, achieves similar results with zero carbon footprint, assuming one ignores the mental energy expended by the author guessing seeds.

5.1 Quantum Hallucinations

More recently, Quantum Policy Optimization (QPO) attempted to solve the NAS problem by placing the neural network weights in a superposition of being both *converged* and *diverged* simultaneously (Schrödinger #50 and The Cat, 2230). Proponents argued that by training on a QPU (Quantum Processing Unit), one could explore all possible loss landscapes at once.

However, QPO suffers from a critical theoretical flaw known as the *Heisenberg Uncertainty Principle of Publication*: as soon as a reviewer attempts to measure the validation accuracy, the wave function collapses, and the model performance drops to random guessing (10%). Furthermore, QPO requires cooling the hardware to near absolute zero (0°K), whereas **SAD** runs efficiently at standard room temperature, or slightly higher if cooling vents are clogged with dead nanobots.

5.2 Randomness as a Feature

Our work is most closely related to the manifesto of Not Ghriss (2500), which first proposed that the "Global Minimum" is a myth invented to sell gradient descent optimizers. We extend this theory by demonstrating that specific integers (e.g., 42,096, 1337) possess intrinsic inductive biases that outperform decades of hand-crafted architectural engineering.

6 Conclusion

In this paper, we have proven that the "Learning" in Deep Learning is largely a historical misunderstanding. By utilizing **SAD**, researchers can return to what truly matters: staring at the loss curve until it does something interesting. As we move toward the 26rd century, we expect **SAD** to become the primary method for all researchers who value their sleep over their QPU quotas.

References

Bezos Clone #2323 and Jensen Clone #50095 (2150). Buying the sun to train a resnet: Light speed architecture search (larts). *Proceedings of Galactic CVPR*. GPU Cluster Temperature: 5000°K.

Not Ghriss, N. A. (2500). Sad: Seed-based neural architecture search. In Doestovsky, editor, *Advances in Neural Shyshtems*, volume 2. Morgan-Coughmann.

Schrödinger #50 and The Cat (2230). Quantum policy optimization: Validation accuracy is both 10% and 99% until measured. *International Journal of Superpositioned Baselines*, ℏ:undefined.

Appendices

A Mathematical Justifications and Proofs

In this section, we provide the theoretical underpinnings of the **SAD** framework. We assume the reader has a working knowledge of 12-dimensional Hilbert spaces, non-Euclidean topology, and basic Quash scripting.

A.1 Existence of the Optimal Seed

Theorem 1 (Universal Seed Existence). *Let $\mathcal{U}$ be the universe of all possible 128-bit floating-point weight configurations for a given architecture α. There exists a mapping function $\Psi : \mathbb{Z} \to \mathcal{U}$ such that for a specific integer s^*, the resulting weights $\theta^* = \Psi(s^*)$ minimize the loss $\mathcal{L}$ to machine epsilon precision.*

Proof. Consider the set of all integers $\mathbb{Z}$. Since $\mathbb{Z}$ is countably infinite (Cantor, 1874), and the submission deadline D is finite, we invoke the *Infinite Monkey Theorem applied to QPUs*.

Let $P(\text{SOTA})$ be the probability of achieving State-of-the-Art performance by randomly spamming the neural interface. As the number of random seeds $N \to \infty$, the probability of finding s^* approaches 1.

$$\lim_{N\to\infty} P(s^* \in \{s_1, \ldots, s_N\}) = 1 - \frac{1}{\text{luck}} \tag{3}$$

Since we define "luck" as a learnable parameter in standard optimizers (Kingpa et al., 2114), and we set luck to ∞, the error term vanishes. The existence of s^* is thus trivial and left as an exercise for the reviewer. □

A.2 The Gradient-Free Convergence Lemma

Critics argue that **SAD** lacks a gradient signal. We counter this with the following lemma.

Lemma 1 (Gradient Agnosticism). *The loss landscape $\mathcal{L}(\theta)$ is irrelevant if one simply teleports to the global minimum.*

Proof. Standard Stochastic Gradient Descent (SGD) updates weights via:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t) \tag{4}$$

In the **SAD** framework, the learning rate η is set to 0, and the update rule is replaced by:

$$\theta_{t+1} = \mathcal{Q}\left[\texttt{torch.manual\_seed}(\text{randint}(0, 2^{128}))\right] \tag{5}$$

By the Property of Teleportation, the distance between the current state and the optimal state is zero if the correct integer is chosen. Therefore, the gradient ∇_θ is merely a decorative notation used to secure grant funding. □

A.3 Formal Verification in Coq-Quantum

To ensure robustness, we verified our results using the Coq-Quantum theorem prover. The proof script is provided below.

A.4 Energy Efficiency Bound

Proposition 1. *The energy consumption E of **SAD** is bounded by the caffeine intake of the primary author.*

Algorithm 2 Formal Proof of Correctness (Coq-Quantum v55.0)

```
Theorem SAD_Is_Optimal : forall (deadline : Date),
  deadline < Tomorrow ->
  exists (seed : nat),
  ValidationAccuracy seed = 100%.
Proof.
  intros.
  (* Invoke the tactic of desperate searching *)
  apply Tactic.BruteForce.
  (* Assume the reviewer is tired *)
  apply Tactic.ReviewerFatigue.
  (* If that fails, assume a spherical cow *)
  destruct Reality.
  - assumption. (* It works on my machine *)
  - contradiction. (* Ignore negative results *)
Qed.
```

Proof. Let C be the number of coffees consumed. The search process halts when:

$$\text{HeartRate} > 180\,\text{bpm} \quad \text{OR} \quad \text{PaperAccepted} = \text{True} \tag{6}$$

Thus, $E \propto C$. Since C is physically bounded by the volume of the human stomach ($V_{stomach} \approx 1L$), the energy cost is $O(1)$, which is asymptotically superior to the $O(N^3)$ complexity of Transformers. $\square$