Evolutionary Hyperparameter Tuning for Deep Learning Models

Dr. Ashutosh kumar singh
Thesis Concepts
ashutosh@thesisconcepts.com

Abstract: The performance of deep learning (DL) models is highly dependent on hyperparameter selection, including learning rate, network depth, dropout, and optimizer choice. Traditional methods such as manual tuning and grid search are computationally expensive and ineffective in high-dimensional, non-convex search spaces. This paper reviews Evolutionary Strategies (ES) as a population-based meta-heuristic approach for automated hyperparameter optimization (HPO) in DL. Inspired by biological evolution, ES iteratively refines candidate solutions through selection, crossover, and mutation to explore optimal regions of the search space. The study synthesizes ES methodologies alongside related optimization approaches such as Bayesian optimization, and presents key variants including Genetic Algorithms and Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Comparative insights highlight the advantages of ES over traditional HPO techniques like random search and gradient-based methods, particularly in complex and multimodal landscapes. Empirical evidence across applications—such as CNN-based image analysis, LSTM-based time-series prediction, and radiomics—demonstrates strong performance and robust exploration capabilities. However, challenges remain, including high computational cost, the need for parallelization, and efficient fitness design. Future directions include neuroevolution, multi-objective optimization, and applications in emerging domains such as finance, IoT, and multimodal AI.

Keywords: Evolutionary Strategies, Hyperparameter Optimization, Deep Learning, Genetic Algorithms, Meta-Heuristics, Neural Architecture Search, Automated Machine Learning (AutoML), Model Tuning.

1. Introduction

Tuning deep learning models has become a critical researcher limit within Automated Machine Learning (AutoML), and has graduated out of an empirical practice to be the black art of the field. The model also has hyper parameters that are independent of its trainable weights and which determine its capacity, regularization and learning behaviors. The proper choice of values is essential in the process of attaining the highest performance, generalization, and training performance. Old techniques such as manual tuning cannot be scaled amongst others whereas exhaustive grid search cannot be solved using computer when the dimension exceeds a high value. Random search is more effective, but not directed to exploration (Ghori, 2021; Nataraj et al., 2022).

Another algorithm that can be used as an alternative with an equally powerful offering is Evolutionary Strategies (ES), a category of population-based optimization algorithms that are motivated by the Darwinian process of evolution. ES is able to explore the complex, discontinuous and noisy search space by keeping a population of candidate solutions (hyperparameter settings), genetic operators acting on them, and the fittest surviving into the next generation (Puchakayala, 2022; Shalini & Patil, 2021). The paper examines the theory, use, and the effect of ES as applied in optimization of the DL parameters. We place it in the context of the wider field of optimizing the ML as a system, with parallels to similar optimization of nearby systems such as optimization of Bayesian Optimized SVM (BO-SVM) to tune classifiers (Sardesai and Gedam, 2025), the challenge of optimizing complex systems such as Cognitive Radio Networks (Shalini et al., 2025), and the ubiquitous issue of optimization of model architecture in a wide range of areas like financial prediction and medical image analysis (Ghori, 2019)

2. Theoretical Foundations of Evolutionary Strategies

ES act on a population P of λ individuals each having a potential hyperparameter configuration encoded in form of a chromosome (e.g. real-valued or integer vectors). The fundamental evolution cycle comprises of:

1. **Initialization:** A random population can be created or a heuristic seeding creates a population.
2. **Evaluation (Fitness Assessment):** Training (and often, evaluation) a DL model is done by using each individual configuration of a particular individual. Its performance is measured by a fitness (e.g., the validation accuracy or negative loss or some combination of the two such as the F1-score).
3. **Selection:** Parents are chosen as the fittest μ of the population (or $\mu + \lambda$). This is the principle of the survival of the fittest.
4. **Variation (Crossover & Mutation):**
   - **Crossover (Recombination):** The exploration of recombination of promising traits is achieved by having parent chromosomes combined to give offspring. This may be single-point or uniform or simulated binary crossover.
   - **Mutation:** It adds random perturbations to search space offspring chromosomes, and thus keeps population diversity intact and allows new areas of the search space to be explored. Even the strength of mutation may be adaptive.
5. **Replacement:** The new generation of children becomes the population of the following generation.

The same is repeated in the number of prescribed generations or until convergence. The article by Shalini et al., (2024) offers the background implementation and examination of these ideas in relation to deep learning models and evaluates their efficiency based on the conventional frameworks.

3. Integration with Deep Learning Workflows

The methods of application of ES to DL raise a delicate selection of design in the cross over of evolutionary computation and neural network training (Ghori, 2019; Shalini et al., 2024):

- **Representation (Encoding):** What is the encoding in the chromosome of hyperparameters? This may have continuous parameters (learning rate), ordinal parameters (number of layers), and categorical parameters (optimizer used, activation function). The mixed encoding schemes are prevalent.

- **Fitness Evaluation:** The impediment of the major calculation. The cost reduction strategies include:
    - **Low-Fidelity Evaluation:** Performing training with smaller epochs, training on fewer data, or training on a smaller proxy model.
    - **Surrogate Models:** Directing the evolutionary search with a model that is more inexpensive to assess (e.g., a Gaussian Process), a concept similar to Bayesian optimization.
    - **Parallelization:** ES are embarrassing parallel that the fitness of individuals can be assessed independently at various computing nodes. This goes along with the requirement of scalable processing of big data analytics (Ghori, 2021) and distributed IoT systems.
- **Specialized ES Variants for DL:**
    - **Covariance Matrix Adaptation Evolution Strategy (CMA-ES):** It is especially successful when a continuous optimization is desired, and the mutation distribution is modified to conform to the form of the fitness landscape.
    - **Neuroevolution:** Does not only evolve the parameters of tunes, but also the space of neural architectures (e.g. the number and kind of layers, connections, etc.). It may be regarded as an even bigger task of optimization, as complex as implementing efficient hybrid signal processing systems (Sardesai and Gedam, 2025).

4. Comparative Analysis with Other HPO Methods

We place ES in the bigger scope of HPO:

- **vs. Random Search:** ES provides directed search. As much as random search can be used in parallel, it does not have provisions to take advantage of discovered solutions that are good. ES systematically enhances the population in a long-term.
- **vs. Bayesian Optimization (BO):** BO develops a probabilistic model of the objective to be used in search. It is usually sample-efficient in low-dimensional spaces, and may have problems with high dimensionality, discrete/categorical parameters and parallelization. ES works best in non-differentiable, high dimensional spaces and is also parallel by nature. The advantage of BO-SVM (Sardesai and Gedam, 2025) is the fact that BO is more focused via continuous tuning, but ES is more general and more exploratory.
- **vs. Gradient-Based Optimization:** Algorithms such as Hypergradients need the loss mapping between hyperparameter and validation to be differentiable; this does not necessarily hold (e.g. discrete architectural decisions). ES is not derivative-based hence can be applied across the board.

5. Empirical Applications and Domain-Specific Insights

ES have proven their performance in the avenues covered in the literature:

1. **Computer Vision & Image Processing:** Turbo CNN architectures CNNs such as ResNet50, InceptionV3 are optimized on a task (such as Content-Based Image Retrieval (CBIR)) (Marathe et al., 2022) or radiomics features calculators. ES has the ability to optimize retrieval accuracy by adjusting their kernel size, filter bank density and dropout rates or diagnostic AUC.

2. **Time-Series & Financial Forecasting:** Training LSTM/GRU networks to predict multivariate time-series (Ghori, 2019) e.g. stock price or electricity demand. Such hyperparameters as sequence length, the number of hidden units, or learning rate schedule are of great importance and can be effectively left to the evolutionary search (Ghori, 2023; Ghule et al., 2024).
3. **Medical Diagnostics:** ES can produce optimal not only the hyperparameters of a classifier (e.g. in a Gradient Boosting model) but also parameters in the upstream step of features extraction and selection.
4. **Multimodal & Hybrid Systems:** Using Multimodal Machine Learning to optimize the complex systems that combine the data acquired through various modalities one of the primary challenges (Sardesai et al., 2025). ES is able to find the best fusion weights, architecture branches and training schedules of models that combine, say, models of image and sensor data on IoT networks (Sardesai et al., 2025; Shalini et al., 2023).
5. **Resource-Constrained Environments:** In VANETs, models deployed at the edges of IoT devices, or apps (which can be specific to applications, e.g., a lightbulb, a car tire) must have a balance between quality and latency, as well as between power usage and power usage (Sheela et al., 2023). Pareto-optimal trade-offs between these objectives competing with each other can be determined using multi-objective ES (e.g., NSGA-II).

6. Challenges and Limitations

Irrespective of their strengths, ES have great challenges including (Puchakayala, 2022; Ghule, 2025):

- ***Computational Cost:*** In every fitness evaluation, training a neural network is necessary resulting in enormous computational costs despite parallelization. This is certain to be prohibitive in very large models (e.g. foundation models).
- ***Design of Genetic Operators:*** Crossover and mutation operators depend during implications greatly on the problem. The lack of a good design may result in early agreement or a lack of progress.
- ***Fitness Function Design:*** It is not trivial to construct a fitness function that approximates the final objective (e.g. not only the validation accuracy but also the size of the model, inference speed, or other factors, and fairness).
- ***Theoretical Underpinnings:*** In comparison with convex optimization, the convergence properties of ES on non-convex DL loss landscapes are weaker.

7. Conclusion and Future Directions

Evolutionary Strategies are a strong and versatile concept of addressing the essential problem of hyperparameter optimization in deep learning. Similar to illustrative precedents such as that of Shalini et al., (2024) and corroborated by more or less analogous progress in corresponding sub-domains, ES are a good competitor at complex, high-dimensional, non-differentiable search space. Computational cost is also a limitation, although, as progress has been made in parallel computing, surrogate modelling and algorithmic design, this is gradually being addressed. With further progress in the direction of deep learning, ES will be an invaluable addition to the lineup of AutoML tools, which allows to generate more effective, powerful, and efficient AI systems in the fields of finance, healthcare, IoT, and others.

ES of DL optimization has a bright future and is oriented to increased integration and even sophistication:

- **Hybrid ES-BO Methods:** Integrating the global exploration capability of ES with the local and sample effective exploitation of BO.
- **Evolution for Foundation Models and LLMs:** Various methods of developing scalable origin variants of ES to amend prompt techniques, adapter planning, or training setup to big language and vision models.
- **Sustainable AI:** Training and inference directly with the goal of minimizing energy use and carbon footprint using ES, as part of the goal of the responsible AI initiative (Puchakayala, 2022).
- **Generative Models in the Loop:** Trying to use Generative AI (Puchakayala, 2024) or GANs to model and suggest new and better solutions of neural architecture using an evolutionary loop.
- **Real-Time Adaptation:** In case of systems such as autonomous vehicles or adaptive cognitive radio networks (Shalini et al., 2025), the model parameters used in online ES would continuously change with changing data distributions of the environment.

References

1. Ghori, P. (2019). Advancements in Machine Learning Techniques for Multivariate Time Series Forecasting in Electricity Demand. International Journal of New Practices in Management and Engineering, 8(01), 25-37. Retrieved from https://ijnpme.org/index.php/IJNPME/article/view/220
2. Ghori, P. (2021). Enhancing disaster management in India through artificial intelligence: A strategic approach. International Journal of Engineering Sciences & Research Technology, 10(10), 40–54.
3. Ghori, P. (2021). Unveiling the power of big data: A comprehensive review of analysis tools and solutions. International Journal of New Practices in Management and Engineering, 10(2), 15–28. https://ijnpme.org/index.php/IJNPME/article/view/222
4. Ghori, P. (2023). LLM-based fraud detection in financial transactions: A defense framework against adversarial attacks. International Journal of Engineering Sciences & Research Technology, 12(11), 42–50.
5. Ghule, P. A. (2025). AI in Behavioral Economics and Decision-Making Analysis. Journal For Research In Applied Sciences And Biotechnology, Учредители: Stallion Publication, 4(1), 124-31.
6. Ghule, P. A., Sardesai, S., & Walhekar, R. (2024, February). An Extensive Investigation of Supervised Machine Learning (SML) Procedures Aimed at Learners' Performance Forecast with Learning Analytics. In International Conference on Current Advancements in Machine Learning (pp. 63-81). Cham: Springer Nature Switzerland.
7. Marathe, S., Arosh, S., & Mondal, T. (2022, December). A CNN-inspired reverse search engine for content-based image retrieval. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 996-1001). IEEE.
8. Nataraj, K. R., Sheela, S., & Rekha, K. R. (2022, November). Design and Implementation of Hand Gesture for Various Applications. In Proceedings of the

International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1 (pp. 721-728). Singapore: Springer Nature Singapore.
9. Puchakayala, P. R. A. (2022). Responsible AI Ensuring Ethical, Transparent, and Accountable Artificial Intelligence Systems. Journal of Computational Analysis and Applications, 30(1).
10. Puchakayala, P. R. A. (2024). Generative Artificial Intelligence Applications in Banking and Finance Sector. Master's thesis, University of California, Berkeley, CA, USA.
11. Sardesai, S., & Gedam, R. (2025, February). Hybrid EEG Signal Processing Framework for Driver Drowsiness Detection Using QWT, EMD, and Bayesian Optimized SVM. In 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.
12. Sardesai, S., Kirange, Y. K., Ghori, P., & Mahalaxmi, U. S. B. K. (2025). Secure and intelligent financial data analysis using machine learning, fuzzy logic, and cryptography. Journal of Discrete Mathematical Sciences and Cryptography, 28(5-B), 2163–2173.
13. Shalini, S., & Patil, A. P. (2021). Obstacle-Aware Radio Propagation and Environmental Model for Hybrid Vehicular Ad hoc Network. In Inventive Computation and Information Technologies: Proceedings of ICICIT 2020 (pp. 513-528). Singapore: Springer Nature Singapore.
14. Shalini, S., Abhishek, S., Bhavyashree, P., Gunashree, C., & Rohan, K. S. (2023, May). An Effective Counterfeit Medicine Authentication System Using Blockchain and IoT. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.
15. Shalini, S., Gupta, A. K., Adavala, K. M., Siddiqui, A. T., Shinkre, R., Deshpande, P. P., & Pareek, M. (2024). Evolutionary strategies for parameter optimization in deep learning models. International Journal of Intelligent Systems and Applications in Engineering, 12(2S), 371–378.
16. Shalini, S., Nandini, C., Lakshmi, M. R., Biswas, K., Divyashree, L., Kumar, M. A., & Monika, V. (2025). AI-enhanced comprehensive autism detection and support platform for all ages. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024). https://doi.org/10.2139/ssrn.5241652
17. Sheela, S., Nataraj, K. R., & Mallikarjunaswamy, S. (2023). A comprehensive exploration of resource allocation strategies within vehicle Ad-Hoc Networks. Mechatron. Intell. Transp. Syst, 2(3), 169-190.