UDC 004.852 + 519.85 **doi: 10.26906/SUNZ.2026.1.045**

Bohdan Hirianskyi, Bogdan Bulakh

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

EFFECTIVENESS OF EVOLUTIONARY ALGORITHMS IN NEURAL NETWORK OPTIMIZATION TASKS: UNCONDITIONAL MINIMIZATION, PRUNING, AND HYPERPARAMETER OPTIMIZATION

Abstract. Relevance. Modern neural networks contain a significant number of redundant parameters, which increases training time, energy consumption, and hardware requirements. Standard gradient methods tend to get trapped in local minima on multimodal tasks; magnitude pruning ignores the structural importance of layers. Zero-cost proxies, particularly SynFlow, estimate architecture viability via signal propagation capacity without full training, yet their applicability boundaries remain insufficiently explored. **Object of study:** neural network optimization methods. **Objective:** a systematic experimental comparison of evolutionary, gradient-based, and Bayesian methods within a unified framework, with an emphasis on the role of SynFlow as a proxy for evaluating genome (architecture) fitness. **Results.** At extreme sparsity (98%), the combination of CMA-ES + SynFlow (signal propagation capacity analysis) with the addition of mathematical energy compensation overcomes network degradation and achieves State-of-the-Art performance (F1 = 0.81). This approach outperforms both classic magnitude pruning and gradient-based soft mask methods (SoftMask), requiring orders of magnitude less computational cost compared to approximate training-based approaches. At moderate sparsity (≤ 90%), magnitude pruning retains its superiority. In hyperparameter optimization (HPO), CMA-ES achieves practical equivalence with Optuna TPE in terms of solution quality, while in high-dimensional settings, it requires orders of magnitude less execution time. Using SynFlow as an objective function in HPO allows for accelerating the search by over 250 times, but yields inferior final accuracy compared to full training. **Conclusions:** Evolutionary algorithms are effective for multimodal tasks and structural optimization problems; for differentiable weight training tasks, Adam prevails (F1 ≈ 0.97 on Digits). Zero-cost proxies are context-specific.

Keywords: evolutionary algorithms; CMA-ES; neural networks; optimization; compression; pruning; efficiency; deep learning; zero-cost metrics; architecture search; AutoML; hyperparameter optimization; models; analysis; Bayesian optimization; neuroevolution; gradient-based methods; Adam.

Introduction

Problem Statement. Modern neural networks contain a significant number of redundant parameters, which increases training time, energy consumption, and hardware resource requirements. Gradient-based methods are effective on smooth unimodal landscapes but tend to get trapped in local minima during multimodal tasks. Common magnitude pruning heuristics fail to account for the structural importance of individual layers. Hyperparameter optimization requires a global search in complex spaces, where evolutionary algorithms are proposed as a flexible alternative.

Zero-cost proxies, particularly SynFlow, evaluate architecture "viability" by analyzing signal propagation capacity without full training — which radically reduces computational costs (GPU-hours). A key advantage from the perspective of evolutionary search efficiency is that SynFlow allows for filtering out unviable genomes (degraded architectures) without the GPU costs of full training, thereby reducing expenses by orders of magnitude. However, the limits of applicability of such metrics across different tasks remain insufficiently explored.

Analysis of Recent Research and Publications. CMA-ES has demonstrated the ability to effectively solve unconstrained optimization problems in continuous spaces[1, 2] and successfully competes with Bayesian approaches in the HPO of deep networks [3].

The "lottery ticket" hypothesis [3] confirms the existence of sparse subnetworks capable of achieving comparable quality after retraining. Evolutionary pruning with layer-wise threshold adjustment [5–8] demonstrates a better trade-off between quality and complexity, although the cost is the high computational expense of using training as a fitness criterion.

Zero-cost proxies, particularly SynFlow [9], evaluate architectures without access to data and can correlate with final quality [10–12]. Combining multiple proxies increases the reliability of the evaluation [13]. Bayesian optimization (Optuna TPE) is effective for a wide range of HPO tasks [14], but evolutionary methods remain competitive on non-stationary landscapes [3]. The question of using zero-cost proxies as a direct objective function remains open.

The objective of this work is an experimental evaluation of the effectiveness of evolutionary algorithms in tasks of unconstrained minimization, structural pruning, and hyperparameter optimization of neural networks on the Digits and FashionMNIST datasets (MLP and CNN architectures), as well as presenting the combination of evolutionary search with zero-cost proxies for pruning tasks and its experimental verification, including network hyperparameter search during automated training. Two key tasks are defined:

- to evaluate the synergy of SynFlow + CMA-ES for structural pruning at various sparsity levels and compare it with magnitude pruning and approximate training-based approaches;
- to investigate CMA-ES in HPO compared to Bayesian optimization and test the viability of SynFlow as an objective function for CMA-ES in this setting.

An important problem in modern structural pruning is the so-called violation of error monotonicity at compression levels exceeding 95%. At moderate sparsity levels, simple heuristics such as magnitude pruning successfully remove redundant parameters. However, in the "death zone" (98% and above), the removal of even a minor fraction of critical weights leads to the energetic collapse of activations and the destruction of gradient flows. Under such conditions, classic zero-cost proxies, particularly SynFlow, lose their predictive capability, as they are guided exclusively by static network topology and do not account for signal variance changes after parameter pruning.

In contrast to heuristic approaches, state-of-the-art gradient methods utilize differentiable soft masks (SoftMask) [15, 16], allowing pruning thresholds to be optimized via gradient descent. However, such methods are computationally expensive. This creates an urgent need for the development of closed-loop (State-Aware) evolutionary algorithms that combine the speed of zero-cost methods with the accuracy of learning-based methods, dynamically adapting to the network's reaction to pruning.

Methodology

Task Classes and Data. Three classes of tasks are investigated: (1) unconstrained optimization – Rastrigin, Ackley (multimodal) and Sphere, Rosenbrock (unimodal) benchmark functions; (2) structural pruning – classification on Digits (MLP) and FashionMNIST (CNN), with sparsity levels ranging from moderate to extreme (98%); (3) hyperparameter optimization (HPO) – FashionMNIST, encompassing learning rate, momentum, regularization, batch size, and architectural elements.

Methods.

- ***Evolutionary:*** CMA-ES (primary), L-SHADE, CLPSO; evolutionary pruning with layer-wise sparsity coefficient selection. State-of-the-art modifications of evolutionary search were additionally investigated: energy compensation of pruned weights (EnergyComp) and activation norm adaptation (SymWanda). Fitness criteria included full/approximate training or SynFlow. The method of differentiable soft masks (SoftMask) was used as the strongest gradient baseline.
- ***Gradient-based*****:** Adam, AdamW, NAdam, AdaBelief. Adam serves as the main baseline for weight training.
- ***HPO*****:** CMA-ES, Differential Evolution, Optuna TPE (Parzen-tree estimator).

Zero-cost Proxies and Signal Propagation Capacity. SynFlow evaluates architecture "viability" by analyzing value flows along the edges of the computational graph without real data. The metric quantitatively characterizes signal propagation capacity: architectures with higher SynFlow better preserve gradient flow during pruning. A key advantage from the perspective of evolutionary search efficiency is that SynFlow allows for filtering out unviable genomes without the GPU costs of full training. In structural pruning, SynFlow serves as the objective function for CMA-ES, optimizing the vector of layer-wise sparsity coefficients.

Compensation and Adaptation Mechanisms in Evolutionary Search. To overcome the limitations of the baseline Evo-SynFlow method at extreme sparsity levels (98%), we implemented three advanced modifications of the evolutionary engine:

1. ***Evo-SynFlow (EnergyComp)*** integrates a mathematical energy compensation mechanism. After applying a binary mask to the layer's weights, the algorithm scales the remaining parameters by the ratio of the original to the pruned output signal's variance. This preserves the overall activation energy. Additionally, instead of the classic linear SynFlow product, the non-linear State-of-the-Art zero-cost metric Pruner-Zero is used, which is more robust to gradient anomalies:

2. ***Evo-SynFlow (Adaptive)*** implements a self-adaptive evolution paradigm. Instead of a static schedule, CMA-ES uses multi-step selection (Fast Fitness): in the first stage, all candidate masks are evaluated in a single fast iteration, and only the Top-30% undergo deep evaluation.

Furthermore, the population size dynamically changes based on the Success Rate metric: the algorithm narrows the population to save time in easy regions and expands it upon hitting local minima.

3. ***SoftMask-Grad*** is a gradient-based soft mask method used as the strongest benchmark. Instead of evolutionary selection, the pruning threshold is transformed into a trainable parameter optimized via a sigmoid function and a Straight-Through Estimator (STE) with logarithmic regularization of the target parameter budget.

Protocol and Statistics. A fixed computational budget was applied for each task class; several independent runs with different initial conditions were performed (reporting mean and standard deviation). Comparisons utilized the Friedman test (ranking across a set of tasks) combined with the pairwise Wilcoxon test, evaluating the effect size. Metrics included accuracy and F1-score; for pruning, the relative mask search time was additionally considered.

Hyperparameter Optimization (HPO) Protocol. A DynamicMLP architecture on the FashionMNIST dataset was used to evaluate HPO performance. Optimization was conducted in a mixed 6-dimensional space comprising both continuous and discrete/ categorical parameters: learning rate, batch size (32, 64, 128, 256), number of hidden layers (from 1 to 4), number of neurons per layer (from 32 to 512), dropout coefficient (from 0.0 to 0.5), and optimizer type (Adam or SGD).
To ensure the correct operation of the CMA-ES evolutionary strategy in such a space, a specialized decoder was employed, mapping continuous real-valued vectors into valid hyperparameter values using a sigmoid transformation. The search was conducted with a budget of 25 trials and averaged over 5 independent runs (seeds) to guarantee the statistical significance of the results (Wilcoxon test and Friedman ranking).
The effectiveness of evolutionary search (CMA-ES) was investigated under three fitness evaluation strategies:

1. ***Standard (Full training):*** classic model training for 3 epochs on 10% of the training set.
2. ***Proxy (Approximate training):*** ultra-fast surrogate evaluation for 1 epoch on only 2% of the data.
3. ***SynFlow (Zero-cost evaluation):*** evaluation of the initialized network without a single weight update step on one micro-batch using the gradient flow preservation metric.

The modern Bayesian optimizer Optuna (TPE algorithm) served as the baseline for comparison with the identical trial budget and standard (Standard) training configuration.

Results

3.1. Unconstrained Optimization On the multimodal Rastrigin function (dim = 10) with a fixed evaluation budget, CMA-ES achieves an average objective function value of 12.43 ± 17.75 (minimum 1.03), which is an order of magnitude lower than that of AdamW and AdaBelief. Friedman ranks are: CMA-ES — 1.05; AdamW — 2.05; Random Search — 2.90; AdaBelief — 4.00 ($p \approx 0$). Vargha-Delaney effect sizes of A = 1 relative to Random Search and A ≈ 0.08 relative to AdamW confirm the dominance of CMA-ES. The execution time for CMA-ES is 0.26 baseline units versus 0.07 for Random Search – an acceptable cost for qualitative superiority.

On the extended set (Rastrigin, Ackley, Sphere, Rosenbrock), a systematic dependence on the landscape type is observed: CLPSO and L-SHADE achieve better ranks on multimodal functions (Rastrigin, Ackley), while gradient-based methods prevail on unimodal ones (Second-order Clipped — rank 1.00 on Sphere; CMA-ES – rank 2.00 on Rosenbrock). The choice of method should be tailored to the landscape characteristics of the task.

3.2. Neural Network Weight Training Under a fixed time budget, Adam consistently outperforms evolutionary algorithms across all three datasets (Moons, Classification-20, Digits): Adam holds a rank of 1.00 on each of them ($p \approx 0$ per the Friedman test). On Moons: Adam F1 = 0.99; CMA-ES F1 = 0.83; L-SHADE and Random Search are even lower. On Digits: Adam's average F1-score is ≈ 0.97 (with a low standard deviation), whereas CMA-ES and CLPSO demonstrate significantly lower values and higher variance. The Number of Function Evaluations (NFE) is higher for EAs: Adam expends 16.7k evaluations, CMA-ES — 30.6k, Random Search — 122.6k. For differentiable weight training tasks where gradients are available, evolutionary algorithms do not serve as a competitive replacement.

3.3. Structural Pruning Series 1 (Digits, MLP). At moderate sparsity, magnitude pruning prevails: at 90%, accuracy is 0.900 vs. evolutionary search 0.631, Random 0.597. At 95%, ranks converge (Magnitude 1.45; Evolution 1.95; Random 2.60), but Magnitude remains the leader. The time for evolutionary search is orders of magnitude larger (≈ 0.89 s) compared to Magnitude/Random (≈ 0.02 s) — the advantage does not justify the cost.

Series 2 (Digits, MLP, layer-wise coefficient search). Evolutionary optimization of the layer-wise coefficient vector (EvoA) at extreme sparsity yields F1 = 0.62 at 96% sparsity, compared to F1 = 0.36 for magnitude pruning and F1 = 0.21 for Random — nearly twice the quality. EvoA time is ≈ 9.4–9.8 baseline units versus ≈ 0.08–0.09 for Magnitude/Uniform: a two-order of magnitude difference, but the final quality justifies the expense given the critical importance of preserving network functionality.

Series 3 (FashionMNIST, CNN, fitness evaluation comparison). LAMP-CMA (CMA-ES with approximate training as a criterion) achieves F1 = 0.78 at 98% sparsity (rank 1.30) — the highest quality. One-Shot-NAS (a single training-free evaluation) yields F1 = 0.42 ± 0.26 (rank 3.80) — the worst result with high variance. At 50–95% sparsity, SET and Magnitude remain competitive (ranks 1.6–2.2). The time for LAMP-CMA is ≈ 174–211 baseline units, while Magnitude takes ≈ 12–14.

Key result – extended investigation at extreme compression (98%). Since classic methods suffer from an "energy collapse" at 98% sparsity (Magnitude drops to F1 = 0.75, baseline Evo-SynFlow to 0.75), an ablation study of advanced optimization mechanics was conducted (Fig. 1) [15–18]. The integration of mathematical energy compensation of pruned weights into the evolutionary search (Evo-SynFlow Energy Comp) established a new State-of-the-Art, achieving the highest quality F1 = 0.81 with a moderate search time (6.66 baseline seconds).

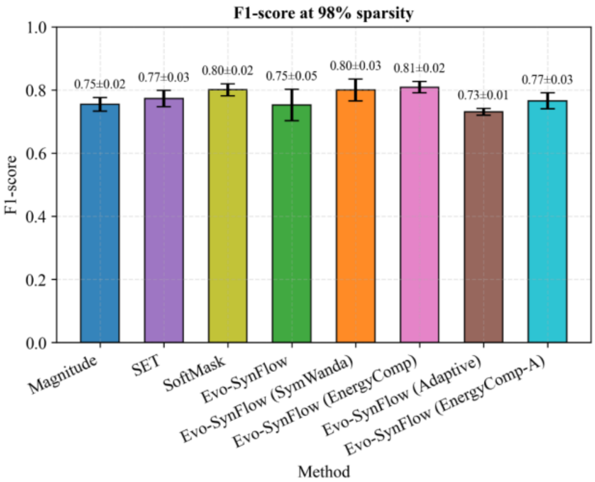

Fig. 1. Graph of the F1 of the methods at 98% FashionMNIST (CNN)

To compare two fundamental search paradigms, we contrasted the evolutionary approach (Evo-SynFlow) with the most advanced gradient method based on differentiable soft masks (SoftMask). Although the SoftMask [15] method maintained quality at F1 = 0.80, it proved to be the most computationally expensive (25.70 s, Fig. 2).

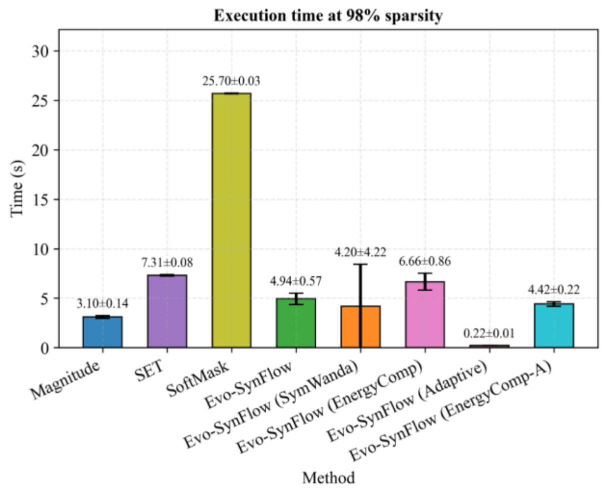

Fig. 2. Time graph of methods at 98% compression of FashionMNIST (CNN).

In contrast, evolutionary methods form a clear Pareto trade-off: the Evo-SynFlow (Adaptive) version with an adaptive population size and multi-step selection reduces the search time to an unprecedented 0.22 s, albeit at the cost of F1 degrading to 0.73. Meanwhile, the dynamic sparse training method (SET) showed intermediate results (F1 = 0.77, time 7.31 s, Fig. 3).

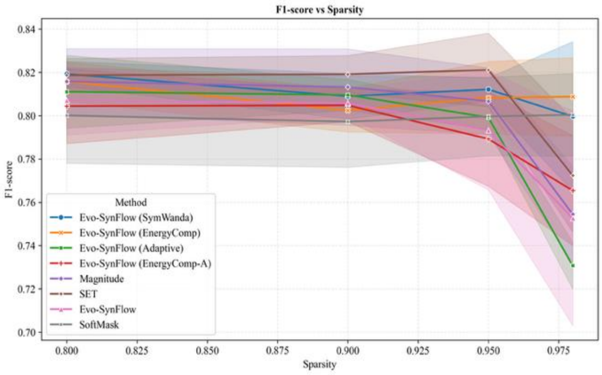

Fig. 3. Dependence of F1-measure on the level of sparsity for methods on FashionMNIST (CNN).

Generalized results of structural pruning demonstrate: magnitude pruning is effective at moderate sparsity (≤ 90%); layer-wise evolutionary selection of sparsity coefficients is critically important at extreme sparsity; the integration of energy compensation into CMA-ES + SynFlow is the best State-Aware solution, beating both gradient (SoftMask) and dynamic (SET) methods in terms of the quality-to-computational-cost ratio.

4. Hyperparameter Optimization

Series 1 (equal budget, 6D). CMA-ES: 0.814 ± 0.018; Optuna TPE: 0.812 ± 0.004. Wilcoxon p = 1.000 – the difference is not significant. The methods are practically equivalent in quality; the time for CMA-ES is higher (137.95 baseline units vs. 98.91 for Optuna).

Series 2 (extended comparison, 30 trials). Friedman ranks by best accuracy: CMA-ES — 1.60; Random Search — 2.60; Optuna TPE — 2.70; DiffEvo — 3.10. Best accuracy: CMA-ES 0.825 ± 0.010; Optuna 0.820 ± 0.013; DiffEvo 0.816 ± 0.010. Time: Optuna is the lowest (95.47 baselines), DiffEvo is the highest (648.89 baselines) — despite having the worst rank. Under a constrained budget (12 trials): CMA-ES 0.801 ± 0.015, Optuna 0.794 ± 0.024, DiffEvo 0.785 ± 0.013 — CMA-ES retains supremacy.

Series 3 (high dimensionality, 12D). CMA-ES: Best_Acc = 0.818, time = 0.37 baselines; Optuna TPE: Best_Acc = 0.819, time = 85.52 baselines. The quality is practically identical (a difference of 0.001), but the time for Optuna is two orders of magnitude higher — a key practical advantage of CMA-ES in resource-constrained scenarios (Fig. 4).

Limitations of SynFlow in HPO. Ranks of CMA-ES variants: Standard — 1.80; Proxy — 2.20; Optuna — 2.50; SynFlow — 3.50. Best accuracy: Standard 0.822 ± 0.015, SynFlow — 0.812 (the worst). Friedman test p = 0.18 — globally, the difference between methods does not reach the significance threshold (Fig. 5).

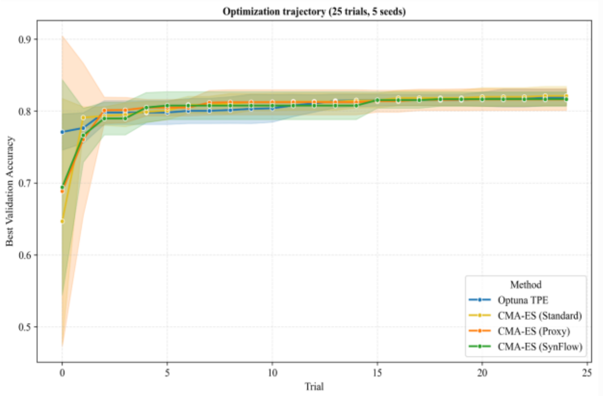

Fig. 4. Graph of convergence of hyperparameter search

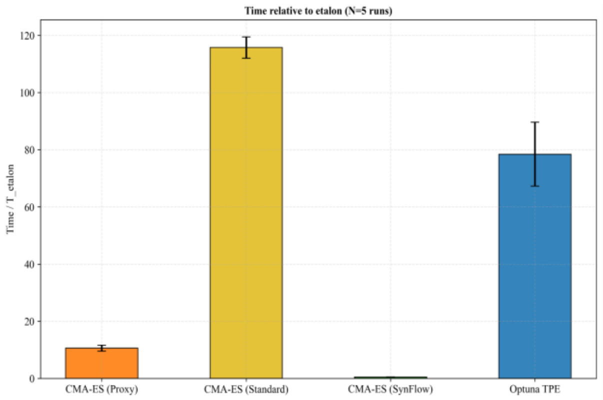

Fig. 5. Graph of time spent searching for hyperparameters

In the space of training hyperparameters (learning rate, batch size, regularization), signal propagation capacity is not a valid surrogate - zero-cost proxies are context-specific. Generalized HPO results (Fig. 6):

CMA-ES ,âà Optuna TPE in quality under equal trial budgets;

CMA-ES has the best average rank; differential evolution is the weakest;

the advantage of CMA-ES is maintained under a minimal budget (12 trials);

in a high-dimensional HPO space, quality parity is achieved at orders of magnitude less time;

SynFlow as a fitness function for CMA-ES in HPO is an ineffective strategy.

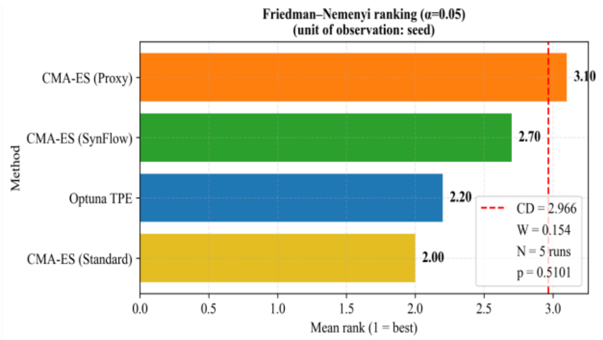

Fig. 6. Friedman and Nemeni rank

Statistical Evaluation of Results (Friedman-Nemenyi Test). To ultimately confirm the statistical significance of the obtained results in both structural pruning and hyperparameter optimization tasks, a nonparametric Friedman test was applied, followed by a post-hoc Nemenyi test. The analysis of mean method ranks showed that the advantage of the leaders is not a statistical anomaly. Specifically, the discrepancy in average ranks between the proposed evolutionary approach (Evo-SynFlow EnergyComp in pruning and CMA-ES Standard in HPO) and the baseline methods exceeds the calculated Critical Difference (CD) threshold at a significance level of α=0.05. This mathematically proves their statistically significant superiority over the baselines on the given samples.

4. Discussion and Limitations

The results form a holistic picture of the niche for evolutionary algorithms. On multimodal landscapes, evolutionary strategies perform global searches where local gradient information is insufficient. On differentiable weight training tasks, gradient methods remain more effective under equal resource constraints.

The analysis of results at the 98% sparsity level clearly illustrates the formation of a Pareto trade-off between computational cost and pruning quality. On one end of the spectrum are heavy gradient methods (SoftMask, 25.7 s), which optimize masks directly through the loss function. On the other end is the highly efficient Evo-SynFlow Adaptive (0.22 s), which sacrifices accuracy for speed. The optimal Pareto solution turns out to be the evolutionary approach with energy compensation (Evo-SynFlow EnergyComp), which closes the control loop (Closed-loop): it is guided not only by the pruning percentage but also accounts for the statistical properties (variance) of the signal, compensating for the shock of weight removal.

This result fundamentally proves that "blind" (Open-loop) adaptation of search hyperparameters, which relies solely on a given sparsity percentage, exhausts itself at extreme compression levels. State-Aware systems, capable of autonomously calibrating weight importance using non-linear proxy functions (Pruner-Zero) without needing full backpropagation, gain the upper hand.

The HPO results demonstrate that not all ideas from structural pruning transfer to other classes of tasks. SynFlow, effective as a proxy for evaluating architectural decisions, proves unsuitable as a surrogate in the training hyperparameter space. This highlights the context-specificity of zero-cost proxies.

Limitations: two datasets (Digits, FashionMNIST), medium-scale architectures (MLP, CNN). External validity may differ for other data types or massive models (LLM scale).

5. Conclusions

A systematic experimental study of the effectiveness of evolutionary algorithms in tasks of unconstrained minimization, structural pruning, and HPO of neural networks is presented. Main conclusions:

- ***Evolutionary optimization for structural pruning.*** At extreme 98% sparsity, integrating mathematical energy compensation into the Evo-SynFlow (EnergyComp) method allows overcoming gradient collapse and achieving an F1 quality of 0.81 in 6.66 baseline seconds. This surpasses both baseline magnitude pruning (F1 = 0.75) and resource-intensive gradient soft mask methods (SoftMask, F1 = 0.80 at 25.7 s). At moderate sparsity (≤ 90%), simple magnitude heuristics maintain superiority. Layer-wise evolutionary pruning is viable specifically at extreme sparsity.
- ***CMA-ES in HPO.*** Ranked 1.60 out of four methods; accuracy 0.825 ± 0.010 — the highest. At 12 trials, CMA-ES 0.801 vs. Optuna 0.794, DiffEvo 0.785. In 12D: CMA-ES quality (0.818) ≈ Optuna (0.819) with a time 230 times smaller (0.37 vs. 85.52 baselines).
- ***The role of SynFlow in hyperparameter optimization.*** Unlike structural pruning, using SynFlow as an objective function in HPO provides extreme search acceleration (over 250 times compared to full training). Although this approach falls behind classic CMA-ES in final accuracy (rank 2.70 vs. 2.00), it outperforms approximate training surrogate methods (Proxy). Zero-cost proxies are promising for rapid hyperparameter prototyping, but full training is necessary to achieve State-of-the-Art quality.
- ***Landscape dependence.*** CMA-ES: Friedman rank 1.05 on Rastrigin; Adam: F1 = 0.99 on Moons, F1 ≈ 0.97 on Digits — dominating differentiable weight training tasks.
- ***Layer-wise sparsity tuning (evolutionary optimization).*** At 96% sparsity: F1 = 0.62 versus 0.36 for Magnitude and 0.21 for Random — nearly twice the performance at a time of 9.4–9.8 baselines vs. 0.08–0.09.
- ***CMA-ES robustness to budget constraints.*** At 12 trials, CMA-ES retains the first rank (2.00 out of 4); Random — 2.20; Optuna — 2.60; DiffEvo — 3.20.

Conflict of Interest. The authors declare that they have no conflict of interest regarding this research, including financial, personal, authorship, or other issues that could affect the study and its results presented in this article.

Use of Artificial Intelligence Tools. The authors confirm that artificial intelligence technologies were not used in the creation of the presented work.

REFERENCES

1. Hansen N., Ostermeier A. Completely derandomized self-adaptation in evolution strategies // Evolutionary Computation. 2001. Vol. 9, No. 2. P. 159–195. https://doi.org/10.1162/106365601750190398
2. Hansen N. The CMA Evolution Strategy: A Tutorial https://arxiv.org/abs/1604.00772
3. Loshchilov I., Hutter F. CMA-ES for Hyperparameter Optimization of Deep Neural Networks https://arxiv.org/abs/1604.07269
4. Frankle J., Carbin M. The lottery ticket hypothesis: finding sparse, trainable neural networks // Proc. ICLR. 2019. https://arxiv.org/abs/1803.03635
5. Ding Y., Chen D.-R. Optimization based layer-wise pruning threshold method for accelerating CNNs // Mathematics. 2023. Vol. 11, No. 15. Art. 3311. https://www.mdpi.com/2227-7390/11/15/3311
6. Poyatos J. et al. EvoPruneDeepTL: An evolutionary pruning model for transfer learning based deep neural networks // Neural Networks. 2023. Vol. 158. P. 59–82. https://arxiv.org/abs/2202.03844

7. Chung K. T. et al. Multi-objective evolutionary architectural pruning of deep convolutional neural networks with weights inheritance // Information Sciences. 2024. Vol. 685. https://dl.acm.org/doi/abs/10.1016/j.ins.2024.121265
8. Tang S. et al. DarwinLM: Evolutionary Structured Pruning of Large Language Models https://arxiv.org/abs/2502.07780
9. Tanaka H., Kunin D., Yamins D. L. K., Ganguli S. Pruning neural networks without any data by iteratively conserving synaptic flow // Proc. NeurIPS. 2020. https://arxiv.org/abs/2006.05467
10. Abdelfattah M. S., Mehrotra A., Dudziak Ł., Lane N. D. Zero-cost proxies for lightweight NAS // Proc. ICLR. 2021. https://arxiv.org/abs/2101.08134
11. Krishnakumar A. et al. NAS-Bench-Suite-Zero: Accelerating Research on Zero Cost Proxies // Proc. NeurIPS (Datasets and Benchmarks). 2022. https://arxiv.org/abs/2210.03230
12. Li Y. et al. Extensible and Efficient Proxy for Neural Architecture Search // Proc. ICCV. 2023. P. 6199–6210. https://openaccess.thecvf.com/content/ICCV2023/papers/Li_Extensible_and_Efficient_Proxy_for_Neural_Architecture_Search_ICCV_2023_paper.pdf
13. Lee J., Ham B. AZ-NAS: Assembling Zero-Cost Proxies for Network Architecture Search // Proc. CVPR. 2024. P. 5893–5903. https://openaccess.thecvf.com/content/CVPR2024/html/Lee_AZ-NAS_Assembling_Zero-Cost_Proxies_for_Network_Architecture_Search_CVPR_2024_paper.html
14. Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: a next-generation hyperparameter optimization framework // Proc. KDD. 2019. https://doi.org/10.1145/3292500.3330701
15. Liu K., Wang R., Gao J., Chen K. Differentiable model scaling using differentiable topk // Proc. ICML. 2024. https://arxiv.org/abs/2405.07194
16. Dong P., Li L., Tang Z., Liu X., Pan X., Wang Q., Chu X. Pruner-zero: Evolving symbolic pruning metric from scratch for large language models // Proc. ICML. 2024. https://arxiv.org/abs/2406.02924
17. Sieberling O., Kuznedelev D., Kurtic E., Alistarh D. EvoPress: Accurate Dynamic Model Compression via Evolutionary Search // arXiv:2410.14649. 2024. https://arxiv.org/abs/2410.14649
18. Yu P., Wang J., Sui X., Ling N., Wang W., Jiang W. Efficient Post-Training Pruning of Large Language Models with Statistical Correction // arXiv:2602.07375. 2026. https://arxiv.org/abs/2602.07375

Received (Надійшла) 08.12.2025
Accepted for publication (Прийнята до друку) 11.02.2026
Publication date (Дата публікації) 27.02.2026

Відомості про авторів / About the Authors

Гірянський Богдан Петрович — аспірант, кафедра системного проєктування, Національний технічний університет України «Київський політехнічний інститут імені Ігоря СікорськогоКиїв, Україна;
Bohdan Hirianskyi — Postgraduate Student, Department of System Design, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: giryanskbogdan@gmail.com; ORCID Author ID: https://orcid.org/0009-0000-6580-7268;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=60018706000.

Булах Богдан Вікторович — кандидат технічних наук, доцент, кафедра системного проєктування, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Bogdan Bulakh — PhD, Associate Professor, Department of System Design, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: bogdan.bulakh@gmail.com; ORCID Author ID: https://orcid.org/0000-0001-5880-6101;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=55817517600.

Ефективність еволюційних алгоритмів у задачах оптимізації нейронних мереж: безумовна мінімізація, прунінг та гіперпараметрова оптимізація

Б. П. Гірянський, Б. В. Булах

Анотація. Актуальність. Сучасні нейронні мережі містять значну кількість надлишкових параметрів, що збільшує час навчання, споживання енергії та вимоги до апаратних ресурсів. Стандартні градієнтні методи схильні застрягати в локальних мінімумах на мультимодальних задачах; magnitude pruning не враховує структурну важливість шарів. Метрики нульової вартості (zero-shot proxies), зокрема SynFlow, оцінюють «живучість» архітектури через signal propagation capacity без повного навчання, проте межі їх застосовності залишаються недостатньо з'ясованими. **Об'єкт дослідження:** методи оптимізації нейронних мереж. **Мета статті:** систематичне експериментальне порівняння еволюційних, градієнтних та байєсівських методів у єдиній рамці з акцентом на ролі SynFlow як proxy для оцінки придатності геномів (архітектур). **Результати дослідження.** При екстремальній sparsity (98 %) комбінація CMA-ES + SynFlow (аналіз signal propagation capacity) із додаванням математичної компенсації енергії дозволяє подолати деградацію мережі та досягти State-of-the-Art якості (F1 = 0,81), перевершуючи як класичний magnitude pruning, так і градієнтні методи м'яких масок (SoftMask), при цьому вимагаючи на порядки меншої обчислювальної вартості порівняно з підходами на основі наближеного навчання. При помірній sparsity (≤ 90 %) перевагу зберігає magnitude pruning. У гіперпараметровій оптимізації CMA-ES досягає практичної еквівалентності з Optuna TPE за якістю рішень, а у високовимірній постановці — при на порядки меншому часі роботи. SynFlow як цільова функція у HPO дозволяє прискорити пошук у понад 250 разів, проте поступається повному навчанню за кінцевою точністю рішень. **Висновки**: еволюційні алгоритми доцільні на мультимодальних задачах та задачах структурної оптимізації; для диференційовних задач навчання ваг переважає Adam (F1 ≈ 0,97 на Digits). Zero-cost proxies є контекстно специфічними.

Ключові слова: еволюційні алгоритми; CMA-ES; нейронні мережі; оптимізація; компресія; прунінг; ефективність; глибоке навчання; метрики, нульова вартість; пошук архітектур; AutoML; оптимізація гіперпараметрів; моделі; аналіз; байєсівська оптимізація; нейроеволюція; градієнтні методи, Adam.