

Exploring Multi-Objective Optimization in Neural Architecture Search: A Critical Review and Problem Formulation

Your Name

April 16, 2025

1 Context and Motivation

Modern neural network architectures are becoming increasingly complex, posing significant challenges in balancing predictive accuracy with constraints such as latency, computational cost, and energy efficiency. Neural Architecture Search (NAS) has emerged as a pivotal AutoML sub-field to automate the discovery of neural network structures. Due to multiple conflicting objectives inherent to NAS, Multi-Objective Optimization (MOO) methods are essential to systematically explore and identify optimal trade-offs.

This seminar project investigates the use and adaptation of MOO techniques within NAS frameworks, with particular focus on evolutionary and gradient-based optimization strategies—two primary approaches currently dominating NAS research, as illustrated in Fig. 1 [2]. While historically Reinforcement Learning (RL) was the pioneering approach, evolutionary and gradient-based methods have increasingly attracted research interest due to their effectiveness and computational efficiency. Hence, we focus exclusively on these two strategies to provide deeper insights into their MOO foundations, adaptations, and innovations within NAS.

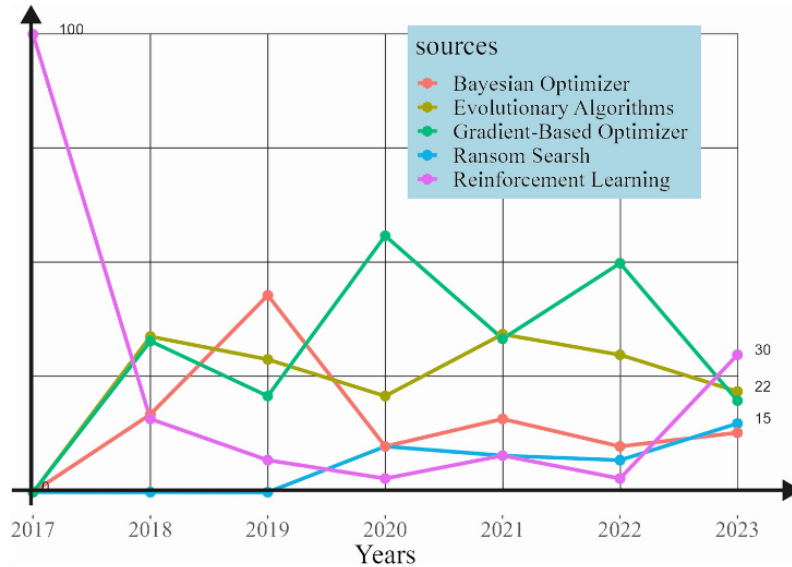


Fig. 8 Percentage of different optimizers used in SSt of NAS algorithms

Figure 1: Popularity trends of NAS search strategies (Credit: [2]).

2 Background

To fully appreciate NAS as an MOO problem, foundational concepts from AutoML, hyperparameter optimization (HPO), NAS, and standard MOO methods are essential.

2.1 Automated Machine Learning and Related Paradigms

Automated Machine Learning (AutoML) aims at automating the end-to-end process of applying machine learning. Prominent subfields include Hyperparameter Optimization (HPO) and Neural Architecture Search (NAS) [1].

2.2 Hyperparameter Optimization (HPO)

HPO methods optimize hyperparameters of a fixed learning algorithm, such as learning rates, regularization terms, or batch sizes. The optimization space is typically low-to-moderate dimensional and comprises continuous or discrete variables.

2.3 Neural Architecture Search (NAS)

NAS automates the selection of the neural network structure itself, covering decisions such as layer types, layer connectivity, and the topology of connections. NAS often involves high-dimensional, combinatorial search spaces, leading to significantly higher complexity and computational costs compared to HPO.

2.4 HPO vs. NAS: A Comparative Summary

Table 1 summarizes key differences between HPO and NAS.

Table 1: Comparison Between HPO and NAS

Aspect	Hyperparameter Optimization (HPO)	Neural Architecture Search (NAS)
Objective	Optimize hyperparameters of fixed model	Discover optimal neural architectures
Search Space	Learning rate, batch size, regularization terms	Layer types, connections, operations, etc.
Complexity	Moderate (continuous & categorical)	High (large, combinatorial, discrete)
Typical Techniques	Bayesian optimization, random search	RL, evolutionary methods, differentiable methods
Evaluation Cost	Lower; fixed model	High; each architecture requires training

2.5 Multi-Objective Optimization (MOO) Concepts

MOO aims to simultaneously optimize multiple conflicting objectives. Key definitions are:

Definition 2.1 (Pareto Dominance). *Given two candidate solutions $x_1, x_2 \in \Omega$, solution x_2 dominates x_1 if $f_m(x_2) \leq f_m(x_1) \forall m$, with at least one strict inequality.*

Definition 2.2 (Pareto Optimality and Pareto Front). *A solution x^* is Pareto optimal if no other solution dominates it. The set of all Pareto-optimal solutions forms the Pareto set, and their image under objective functions forms the Pareto front.*

2.6 Prominent Multi-Objective Algorithms

- **Non-dominated Sorting Genetic Algorithm (NSGA-II)**: evolutionary method using Pareto-based sorting.

- **Non-Dominated Sorting Crisscross Algorithm (NSCA)**: efficient handling of many-objective problems.
- **Multiple Gradient Descent (MGD)**: gradient-based multi-objective optimization method for deep networks.

3 Comprehensive Review and NAS-MOO Problem Formulation

NAS inherently involves multi-objective optimization. Commonly optimized objectives include predictive performance (accuracy, perplexity) and efficiency metrics (latency, energy, complexity).

Formally, NAS as MOO can be defined as:

$$\begin{aligned} \min_x \quad & F(x) = [f_1(x), f_2(x), \dots, f_M(x)]^T, \\ \text{subject to} \quad & x \in \Omega, \quad w^*(x) = \arg \min_w \mathcal{L}_{\text{train}}(x, w), \end{aligned} \tag{1}$$

where x denotes architectural variables, and $w^*(x)$ optimal model weights.

3.1 Challenges in NAS-MOO

NAS-specific challenges include large discrete spaces, noisy evaluations, expensive training, and lack of ground-truth Pareto fronts. These challenges drive innovations and adaptations in standard MOO techniques.

4 Proposed Methodology

We will experimentally compare two key NAS search strategies:

1. **Gradient-based Differentiable NAS (MODNAS)**: utilizing hyper-networks and MGD to approximate Pareto fronts rapidly.
2. **Evolutionary Multi-Objective NAS (NSGA-II)**: a robust population-based evolutionary search.

Additionally, we will employ **Random Search (RS)** as a baseline.

4.1 Gradient-based MODNAS Framework

We adopt a cell-based NAS structure, using continuous relaxation and differentiable architecture parameters (α). MODNAS relies on hyper-networks to predict performance across objectives, effectively applying MGD to balance conflicting gradients and discover architectures across multiple devices and objectives simultaneously.

4.2 Evolutionary NAS Framework (NSGA-II)

NSGA-II maintains a diverse population of architectures, employing Pareto-based selection, crossover, and mutation to progressively identify Pareto-optimal solutions.

5 Experiments and Expected Outcomes

Experiments will be performed on standard NAS benchmarks (NASBench-101, NASBench-201, DARTS, etc.), comparing MODNAS, NSGA-II, and Random Search. Key metrics include hypervolume indicator and diversity of the Pareto-front.

6 Conclusion and Future Work

This project seeks deeper clarity regarding how NAS leverages and contributes to MOO. Through systematic comparison of evolutionary and gradient-based NAS, we aim to identify meaningful insights and potential generalizable contributions to broader MOO theory and methods.

References

- [1] Mitra Baratchi, Can Wang, Steffen Limmer, Jan N. van Rijn, Holger Hoos, Thomas Bäck, and Markus Olhofer. Automated machine learning: Past, present and future. *Artificial Intelligence Review*, 57(122), 2024.
- [2] Sasan Salmani Pour Avval, Nathan D Eskue, Roger M Groves, and Vahid Yaghoubi. Systematic review on neural architecture search. *Artificial Intelligence Review*, 58(3):73, 2025.