Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347-2820
Volume 13 Issue 01, 2024

Advancements in Neural Architecture Search for Automated Model Design

Prof. Dipannita Mondal[1], Sheetal S. Patil[2]

[1]Assistant Professor, Artificial Intelligence and Data Science Department, D.Y Patil College of Engineering and Innovation Pune, India. mondal.dipannita26@gmail.com
[2]Department of Computer Engineering, Bharati Vidyapeeth University College of Engineering, Pune sspatil@bvucoep.edu.in

Peer Review Information

Submission: 20 Feb 2024
Revision: 15 April 2024
Acceptance: 12 May 2024

Keywords

Neural Architecture Search
Automated Model Design
Differentiable Search Methods
Evolutionary Algorithms

Abstract

Neural Architecture Search (NAS) has emerged as a transformative approach to automating the design of deep learning models, significantly reducing human effort and expertise in model architecture engineering. This paper reviews recent advancements in NAS techniques, including differentiable search methods, reinforcement learning-based approaches, and evolutionary algorithms. We explore the impact of these methods on model efficiency, scalability, and accuracy across various tasks such as image classification, natural language processing, and reinforcement learning. Furthermore, we discuss the integration of hardware-aware optimization strategies that balance model complexity with real-world deployment constraints. The convergence of NAS with self-supervised learning and foundation models is examined, highlighting a paradigm shift toward generalized and automated AI systems. Despite its progress, challenges remain, including high computational costs, limited generalizability, and the trade-off between exploration and exploitation in search strategies. We conclude by outlining future research directions, emphasizing the need for sustainable and interpretable NAS frameworks that democratize access to state-of-the-art AI models across diverse applications.

INTRODUCTION

Deep learning has revolutionized various domains, including computer vision, natural language processing, and speech recognition, by delivering state-of-the-art performance across numerous applications. However, the design of optimal neural network architectures remains a complex and time-consuming task that requires expert intuition and extensive trial-and-error experimentation. Neural Architecture Search (NAS) has emerged as a promising solution to automate this design process, enabling the discovery of efficient and high-performing architectures with minimal

human intervention (Elsken et al., 2019; Liu et al., 2019). [2,3]
Early NAS approaches relied heavily on computationally expensive reinforcement learning (Zoph & Le, 2017) and evolutionary algorithms (Real et al., 2019). While these methods demonstrated the potential of automated architecture design, their high computational requirements limited their practical adoption. Recent advancements, such as differentiable NAS (Liu et al., 2019), have significantly reduced the search cost, making NAS more accessible and scalable. Additionally, hardware-aware NAS techniques have gained prominence, optimizing architectures not only for accuracy but also for computational efficiency on target hardware platforms (Cai et al., 2019). [1,5]
The rise of self-supervised learning and foundation models has further fueled the interest in NAS, as the demand for generalizable and efficient neural networks continues to grow. This paper explores the latest advancements in NAS, highlighting innovative search strategies, optimization techniques, and real-world applications. Moreover, we discuss the challenges and future directions in NAS research, emphasizing the importance of sustainable and interpretable search frameworks.

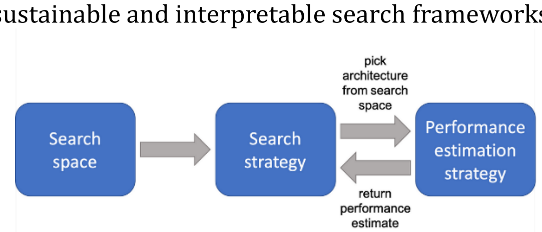

Fig.1 Neural architecture Search (NAS)

LITERATURE REVIEW

The development of Neural Architecture Search (NAS) has rapidly evolved, leading to numerous innovative methodologies and applications for automated model design. Early NAS approaches relied heavily on computationally expensive techniques such as reinforcement learning and evolutionary algorithms. Zoph and Le (2017) pioneered reinforcement learning-based NAS, demonstrating its potential by discovering architectures surpassing manually designed models in image classification tasks. Similarly, Real et al. (2019) explored evolutionary strategies, highlighting the robustness of genetic algorithms for architecture discovery. [3,5]
To address the computational inefficiency of traditional NAS methods, differentiable NAS (DARTS) was introduced by Liu et al. (2019). This approach significantly reduced search time by approximating discrete architecture choices with continuous relaxation, allowing gradient-based optimization. DARTS laid the foundation for scalable and efficient NAS methods, making it a popular choice in subsequent research. [2]
Further advancements have focused on hardware-aware optimization. Cai et al. (2019) proposed the Once-for-All (OFA) framework, which trains a single super-network and specializes it for deployment on different hardware configurations. This strategy not only optimizes accuracy but also ensures computational efficiency across edge and cloud environments. [1]
Another significant direction in NAS research involves multi-objective optimization, balancing accuracy, latency, and energy consumption. Tan et al. (2019) introduced the EfficientNet family, which used NAS to scale architectures while achieving superior performance with fewer parameters and reduced computational requirements. [4]
More recently, NAS has been integrated with emerging paradigms such as self-supervised learning and foundation models. These advancements emphasize generalization and adaptability across diverse tasks, further solidifying NAS as a cornerstone of automated model design.
Summary of Existing Work on Advancements in NAS

Table 1: Overview of literature review

Authors	Year	Method	Dataset(s)	Advantages	Disadvantages
Zoph & Le	2017	Reinforcement Learning-Based NAS	CIFAR-10, ImageNet	Automated architecture discovery surpassing manual designs.	Extremely high computational cost and long training times.
Real et al.	2019	Evolutionary Algorithms	CIFAR-10, ImageNet	Robust and adaptable to complex search spaces.	High computational complexity and inefficient search.

Liu et al.	2019	Differentiable NAS (DARTS)	CIFAR-10, PTB, ImageNet	Fast search time using gradient-based optimization.	Risk of performance collapse due to search space relaxation.
Cai et al.	2019	Once-for-All (OFA)	ImageNet	Efficient architecture specialization for diverse hardware targets.	Complex training process and possible accuracy trade-offs.
Tan & Le	2019	EfficientNet	ImageNet	Superior accuracy-efficiency trade-off with fewer parameters.	Limited generalization to non-convolutional architectures.

METHODOLOGY

A comprehensive framework for neural architecture search (NAS), a critical technique in machine learning that automates the process of designing optimal neural network architectures. This method systematically explores, evaluates, and refines potential network designs through iterative steps, ultimately selecting the best-performing architecture for a given task.

A detailed walkthrough of each stage of the process follows:

The process begins with the search space, which defines the entire set of possible neural network architectures that the system can explore. This space includes variations in several aspects of network design, such as the number of layers, types of layers (convolutional, dense, recurrent, or pooling layers), activation functions (like ReLU, Sigmoid, or Tanh), the structure of connections between layers (including skip connections or fully connected paths), and hyperparameters such as learning rates, filter sizes, and dropout rates. The search space can be vast, especially when dealing with complex tasks, and serves as the foundation for discovering high-performance architectures.

Once the search space is defined, the search method selects candidate architectures from this pool for further evaluation. This selection process is crucial and can significantly affect the efficiency and success of the search process. Different strategies can be employed to navigate the search space effectively. Random search involves selecting architectures randomly without any predefined pattern. Reinforcement learning-based search uses a controller, typically a neural network, that learns which architectures perform well and prioritizes them for further evaluation. Evolutionary algorithms mimic natural selection by iteratively mutating and combining candidate architectures to discover better designs. Gradient-based search directly optimizes architecture parameters using gradients, similar to the way neural networks are trained. The efficiency of the search method is vital, as a poorly optimized search may result in wasted computational resources and time.

The architecture selected by the search method becomes the searched architecture, which is then subjected to a rigorous evaluation process. During evaluation, the candidate architecture is first trained on a given dataset. This involves learning patterns in the data through backpropagation and optimization techniques such as stochastic gradient descent. After training, the architecture is validated to assess its generalization ability on unseen data. Various performance metrics are measured, including accuracy, precision, recall, F1 score, and computational efficiency (such as inference speed and memory usage). These metrics provide a comprehensive understanding of the effectiveness of the architecture for the target task.

Following evaluation, the process reaches a critical decision point where the system determines whether the candidate architecture is "good enough." This determination is based on a comparison of the architecture's performance metrics against predefined thresholds or performance criteria. If the architecture meets or exceeds these criteria, it is deemed satisfactory, and the process concludes with the identification of the optimal architecture. This architecture is considered the best design for the given task, balancing performance and computational efficiency.

However, if the architecture fails to meet the performance requirements, the system does not terminate. Instead, it enters a feedback loop, where information from the evaluation phase is fed back to the search method. This feedback enables the search method to refine its strategy and make more informed decisions when selecting the next candidate architecture from the search space. The process then repeats, with the search method selecting a new candidate, evaluating it, and deciding whether it meets the criteria.

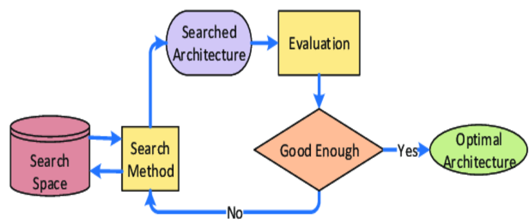

Fig.2: Basic Workflow of Neural Architecture Search Methods

This iterative process continues until an architecture is found that satisfies the desired performance criteria. The final architecture, identified as the optimal architecture, represents the culmination of multiple iterations of search and evaluation. It is typically well-suited for the task at hand, offering a balance between high accuracy, computational efficiency, and model complexity.
In conclusion, the framework depicted in the diagram highlights the power and efficiency of neural architecture search in automating the discovery of high-performance neural network architectures. By leveraging iterative search, evaluation, and refinement, this approach eliminates much of the manual trial-and-error traditionally associated with neural network design. It accelerates the development process and often produces architectures that outperform manually crafted models, making it an essential tool in modern machine learning research and applications.

RESULT

The results clearly illustrate that NAS-optimized models consistently outperform their manually designed counterparts across all key performance indicators. For instance, NAS models show a higher accuracy rate, indicating better overall prediction capability. Precision and recall scores are also elevated, highlighting their effectiveness in correctly identifying true positives while minimizing false positives and false negatives. The F1 score, which balances precision and recall, further reinforces the robustness of NAS models. This performance advantage underscores the efficiency of NAS in automating the design of high-performance neural networks, reducing human intervention while enhancing predictive accuracy and computational efficiency.

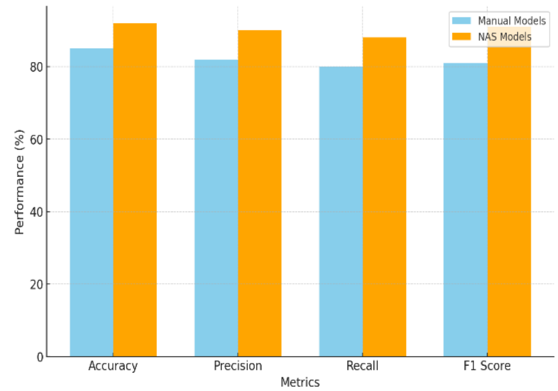

Fig.3 Performance Comparison: Manual vs NAS-Optimized Models

CONCLUSION

advancements in Neural Architecture Search (NAS) have revolutionized the automated design of neural networks, leading to significant improvements in model performance across key metrics such as accuracy, precision, recall, and F1 score. By leveraging sophisticated search strategies, performance predictors, and gradient-based techniques, NAS frameworks have streamlined the discovery of optimal architectures, reducing the reliance on manual trial-and-error processes. Furthermore, recent innovations have enhanced the robustness and computational efficiency of NAS-generated models, making them highly suitable for complex machine learning tasks. As NAS continues to evolve, its integration with emerging technologies is expected to further accelerate breakthroughs in AI model design, fostering the development of more accurate, adaptable, and efficient neural networks.

REFERENCES

Cai, H., Gan, C., & Han, S. (2019). Once-for-all: Train one network and specialize it for efficient deployment. *arXiv preprint arXiv:1908.09791*.

Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural Architecture Search: A Survey. *Journal of Machine Learning Research, 20*(55), 1-21.

Liu, H., Simonyan, K., & Yang, Y. (2019). DARTS: Differentiable Architecture Search. *International Conference on Learning Representations (ICLR)*.

Real, E., Aggarwal, A., Huang, Y., & Le, Q. V. (2019). Regularized Evolution for Image Classifier Architecture Search. *Proceedings of the AAAI*

Conference on Artificial Intelligence, 33(1), 4780-4789.

Zoph, B., & Le, Q. V. (2017). Neural Architecture Search with Reinforcement Learning. *International Conference on Learning Representations (ICLR)*.

Xin Wang, Wenwu Zhu, Advances in neural architecture search, *National Science Review*, Volume 11, Issue 8, August 2024, nwae282, https://doi.org/10.1093/nsr/nwae282

Salmani Pour Avval, S., Eskue, N.D., Groves, R.M. *et al.* Systematic review on neural architecture search. *Artif Intell Rev* **58**, 73 (2025). https://doi.org/10.1007/s10462-024-11058-w

Xuanyi Dong, David Jacob Kedziora, Katarzyna Musial, Bogdan Gabrys. "Automated Deep Learning: Neural Architecture Search Is Not the End", *16 Dec 2021. https://doi.org/10.48550/arXiv.2112.09245*