Surrogate-Assisted Diversity Estimation in Neural Ensemble Search

Udeneev Alexandr Vladimirovich

Scientific Advisor: Ph.D. in Physics and Mathematics, Oleg Yurievich Bahteev

Moscow Institute of Physics and Technology
My First Research Paper

April 17, 2025

Goals

Research Goal:

Reduce the time spent on building an effective ensemble.

Task:

Describe transformation of the DARTS search space of neural network architectures. Describe the surrogate function and its application in ensemble construction.

Introduction

The neural network architecture space $\mathcal{A}$ is exceedingly large (on the order of 10^{24} architectures in our case), raising the question of developing an efficient method for searching for the optimal ensemble $S \subset \mathcal{A}$, i.e., the ensemble that achieves the highest accuracy.

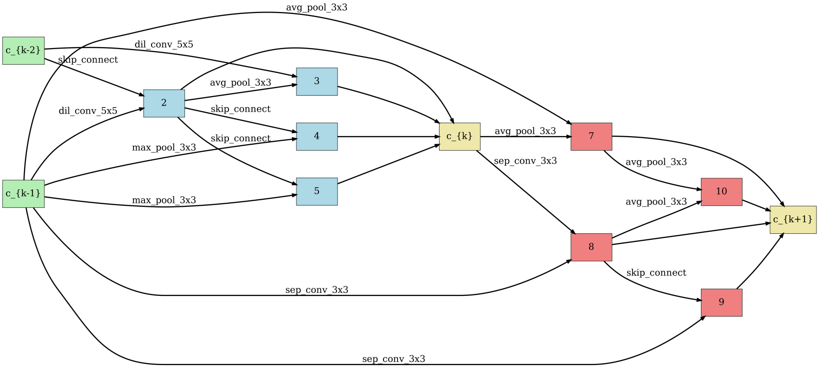


Рис.: Example of one architecture in NAS-Bench-201 format

Literature

Our work is based on the following articles:

- Similarity surrogate-assisted evolutionary neural architecture search with dual encoding strategy (Xue, Zhang, Neri, 2024)
- Neural Predictor for Neural Architecture Search (Wen et al., 2019)
- Few-shot Neural Architecture Search (Zhao et al., 2020)

Problem statement

The primary objective of NES is to find an optimal ensemble of neural networks whose architectures lie within the NAS search space.

Let denote $\alpha \in \mathcal{A}$ as a network architecture and $\omega(\alpha)$ as its corresponding weights. The action of this network on an input x is denoted by $f_\alpha(x, \omega(\alpha))$. Let $S \subset \mathcal{A}$ be a subset of architectures. Then, the NES problem can be formally described as follows:

$$\min_{\mathcal{S}} \mathcal{L}_{val} \left(\frac{1}{|S|} \sum_{\alpha \in S} f_\alpha(x, \omega^*(\alpha)) \right)$$

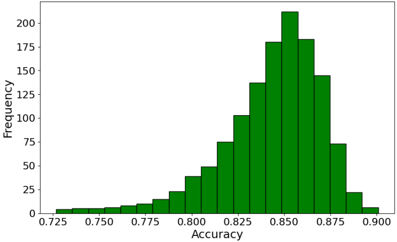
$$\text{s.t.} \quad \forall \alpha \in \mathcal{S} : \; \omega^*(\alpha) = \arg\min_{\omega(\alpha)} \mathcal{L}_{train}(f_\alpha(x, \omega(\alpha)))$$

Thus, in addition to searching over a vast number of architectures, we now also need to find the optimal ensemble composition.

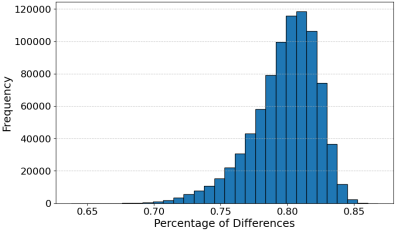
Problem solution

Surrogate Function

We propose a function to estimate model similarity (e.g., using the Jensen-Shannon distance on model predictions from the test dataset) to guide the selection of an optimal ensemble, where "optimal"refers to maximizing the ensemble's overall predictive accuracy.



Distribution of model accuracies



Distribution of model diversity

Problem solution

The algorithm for training the surrogate function is as follows:

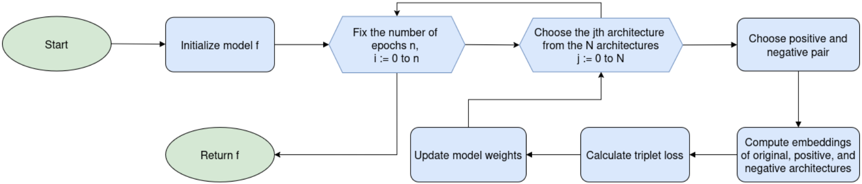


Рис.: Training the surrogate function

Problem solution

Our main algorithm is as follows:

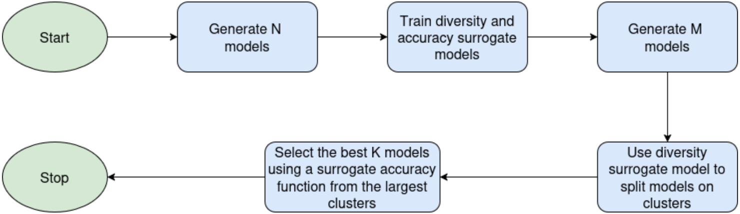


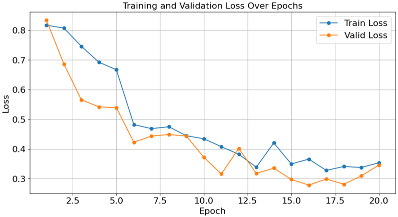
Рис.: The main algorithm of our model.

Computational experiment

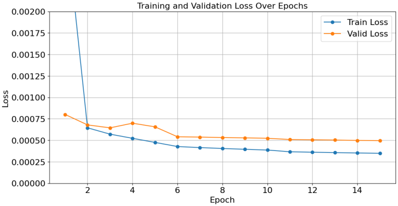
In the base experiment, we generate 250,000 candidate architectures from the DARTS search space. From these, we select the top five architectures based on their validation performance. Each selected architecture is then trained individually from scratch. The average test accuracy achieved by the individual models is approximately XX%, while the ensemble composed of these five models attains a higher accuracy of approximately YY%.
We anticipate that our proposed method will outperform both Deep Ensemble Search and Random Search in terms of final test accuracy.

Error analysis

We employ the Mean Squared Error (MSE) loss to train the accuracy surrogate function and utilize the Triplet Loss to train the diversity surrogate function.
To assess the learned diversity representation, we compute the Pearson correlation between embedding distances and model similarity, obtaining approximately –0.4. This suggests that the surrogate captures their inverse relationship.



Training of diversity function



Training of accuracy function

Datasets

In our work, we use the following data sets:

- CIFAR10/100
- MNIST

Results

1. We propose a method for encoding the DARTS search space into a representation suitable for training a Graph Neural Network (GNN), where graph nodes correspond to operations within the network.
2. We propose a way for training the surrogate function to predict the diversity of architectures.
3. We adapt surrogate functions for ensemble construction, taking into account both predictive performance and architectural diversity.