Towards a Neural Network Evaluation Mechanism for Neural Architecture Search

Atah Nuh Mih[1], Francis Palma[2], Hung Cao[1]

[1]Analytics Everywhere Lab, University of New Brunswick

[2]SE+AI Lab, University of New Brunswick

Abstract

Neural Architecture Search (NAS) automates the design of neural networks but often struggles with flexibility in evaluation criteria, particularly for hardware-aware searches. We introduce a novel neural network evaluation mechanism that converts neural networks into string representations, extracts vector embeddings, and predicts evaluation metrics. Our implementation consisted of a T5 transformer encoder trained to predict accuracy, latency, and memory. Experiments were conducted on a benchmark neural network dataset NATS-Bench consisting of 48,393 architectures and their evaluation metrics. Our method effectively predicts the metrics, showing stronger correlations for predicted vs reported values in Kendall's τ analysis.

Introduction and Motivation

1. Neural architecture search (NAS) automates neural network design but struggles with the evaluation of these nets.
2. A unified method is needed to evaluate performance and hardware cost for all neural networks

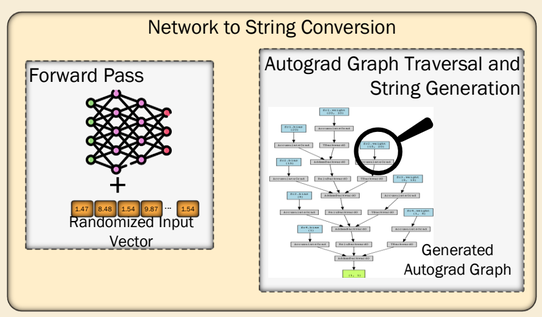
Proposed Method

1. A network-to-string conversion mechanism, making it adaptive to all types of neural architectures.
2. An evaluator consisting of an encoder-predictor network designed to include any evaluation metric(s)
3. A method to evaluate candidate architectures in neural architecture search

Methodology

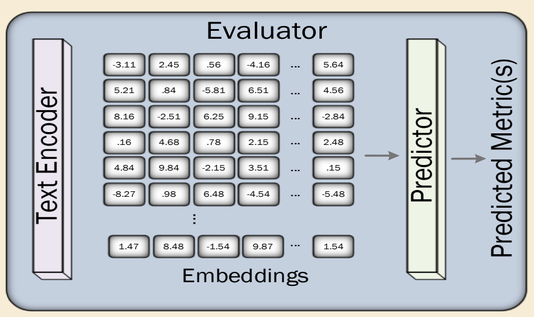
Network to String Conversion

- Traverses computational graph to get network operations
- Converts this operations into string of text



Evaluator

- Consists of an encoder and a prediction head
- Encodes input tokens into a high dimensional representation
- Uses the encoding to predict the metric(s)



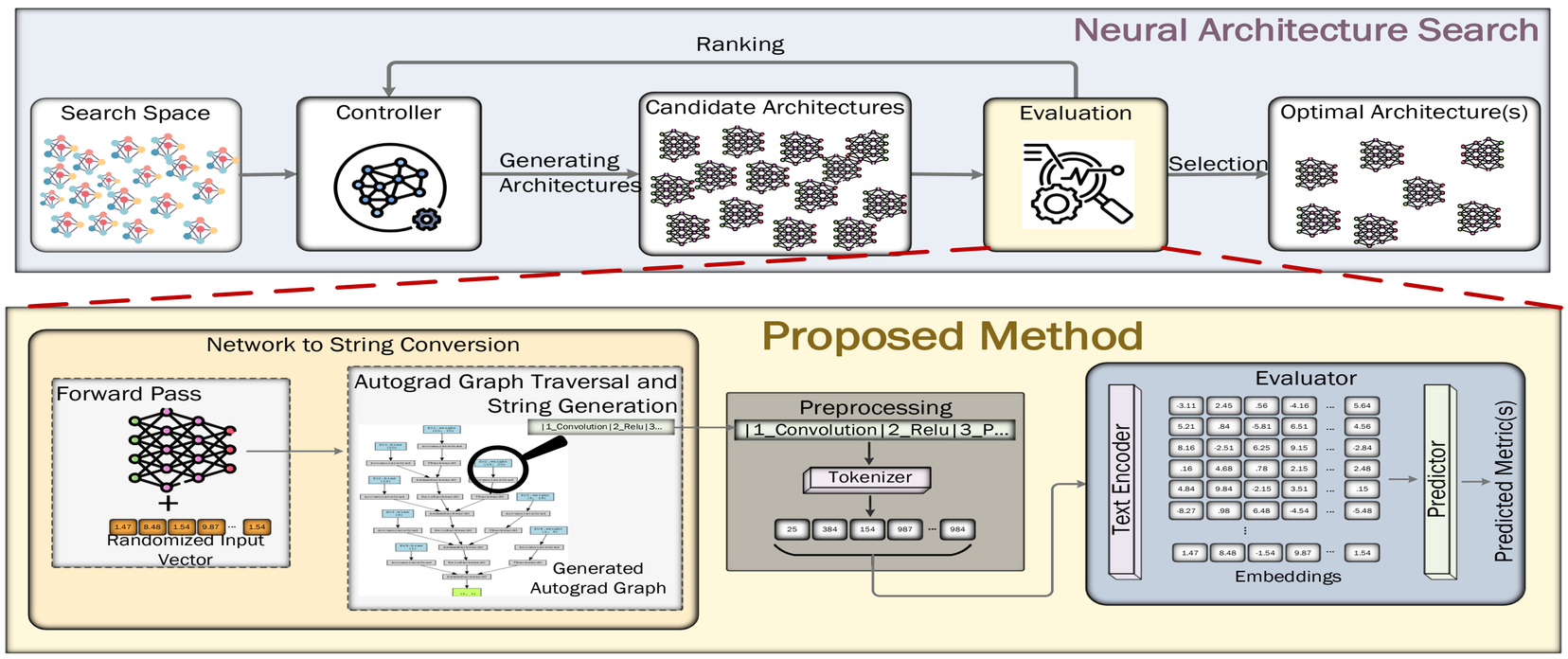


Figure 1. Our proposed method.

Experiments

- Experiments conducted on benchmark dataset NATS-Bench consisting of topology (TSS) and size (SSS) search spaces
- Attributes reported on CIFAR10, CIFAR100, and ImageNet16-120
- Evaluator trained to predict memory and latency in addition to accuracy

NATS-Bench: TSS

- Consists of 15,625 architectures
- Search space varies cell blocks with pre-defined operations

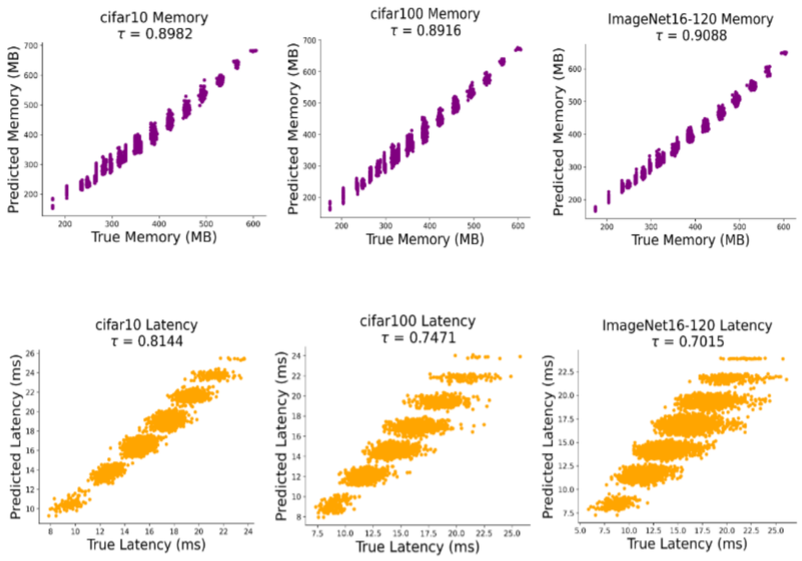


Figure 2. Correlation for predicted and reported memory and accuracy in TSS

NATS-Bench: SSS

- Consists of 32,768 architectures
- Search space varies number of channels in layers

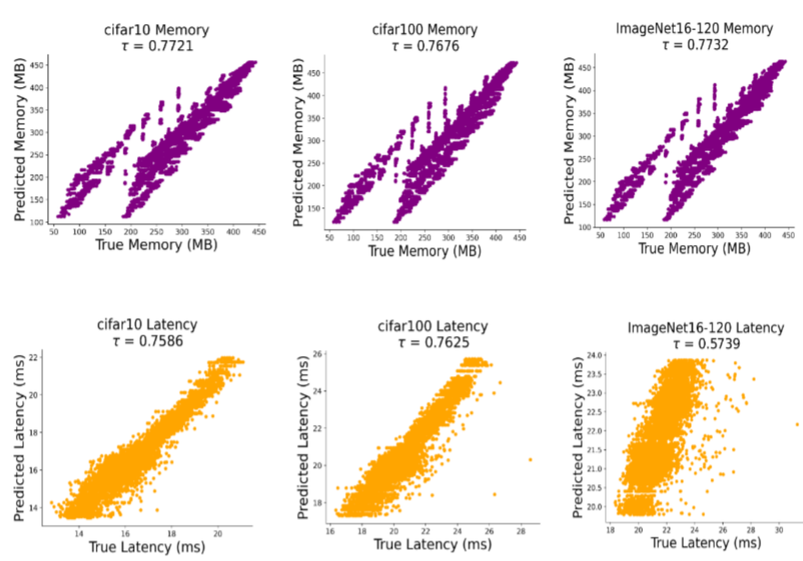


Figure 3. Correlation for predicted and reported memory and accuracy in SSS

Implementation

We used the following setup for our implementation

1. Evaluator: T5 transformer encoder and single layer prediction head
2. Metrics: Accuracy, Latency, and/or Memory
3. Hardware Config: NVIDIA GeForce RTX 4090 Server

Conclusion and Future Work

1. We proposed a method for evaluating the performance of neural architectures
2. Tested our approach on benchmark NATS-Bench and found high correlation for predicted metrics
3. Future work aims to design a lightweight evaluator

This work was supported by the UNB University Research Fund (URF-NF-EXP 2024-16)

atah.mih@unb.ca