2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Proposal of a new zero-shot evaluation index for simple CNN

Chisato Takahashi† Kenya Jin'no†

†Department of Intelligent Systems, Faculty of Knowledge Engineering, Tokyo City University
1-28-1 Tamazutumi, Setagaya, Tokyo 158-8557, Japan
Email: g1923055@tcu.ac.jp, kjinno@tcu.ac.jp

Abstract—Network Architecture Search (NAS), which aims to optimize the structure of neural networks themselves, has attracted much attention in recent years. The evaluation of the structure of a neural network in NAS is basically performed by actually training the neural network and measuring its performance. However, this method requires an enormous amount of computation. For this reason, the other zero-shot method that evaluates the structure without actually performing the training has begun to be proposed. The ultimate goal of this research is to create an evaluation index that can evaluate the structure in a zero-shot manner for NAS. In this article, we experimentally investigate the relationship between basic CNN structures and their performance, we create an index that can measure performance in a zero-shot environment.

1. Introduction

In recent years, the processing power of neural networks has dramatically improved, and remarkable results have been reported in such fields as image classification and natural language processing. These neural network designs are based on the high-level expertise of the researcher's previous experience. In contrast, recently, automated machine learning (AutoML)[1][2], which aims to automate everything from data collection and processing to feature design, neural network generation, and neural network operation, has attracted much attention[1][2]. One type of AutoML is Network Architecture Search (NAS)[1][2], which searches the structure of a neural network itself.

Until now, the evaluation of the structure of a neural networks in NAS is based on actually learning and measuring its performance. This requires an enormous amount of computation time to train for the type of networks to be explored. The first NAS proposed using reinforcement learning required 28 days of computation time on a machine with 800 GPUs connected in parallel[3]. In order to reduce the computation time of NAS, research has begun to propose evaluating the structure of the network without actually performing the training. This is called as zero-shot method[4][5] that evaluates the network structure without learning. Specifically, it obtains an index for each structure based on the results of theoretical analysis of the neural network's expressive power, which is the function reproduction power that the neural network possesses. If this index can correctly indicate the ability that the structure of the neural network has, then it is possible to dramatically improve the search time of NAS because the execution of learning is no longer necessary. However, it is not clear whether this index is sufficient to search a relatively small structured network, because the dependence of the performance on the initial value is large even if the structure is the same.

Against this background, this paper aims to experimentally investigate the relationship between the basic structure and performance of CNNs and to propose a more appropriate zero-shot evaluation metric.

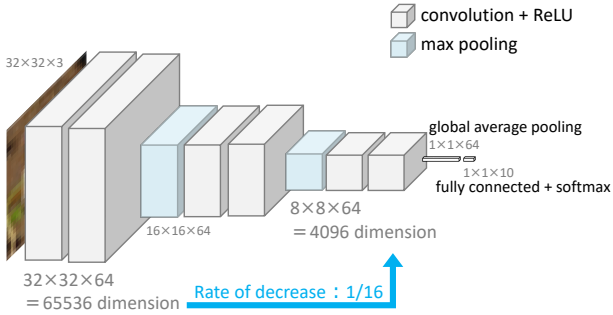


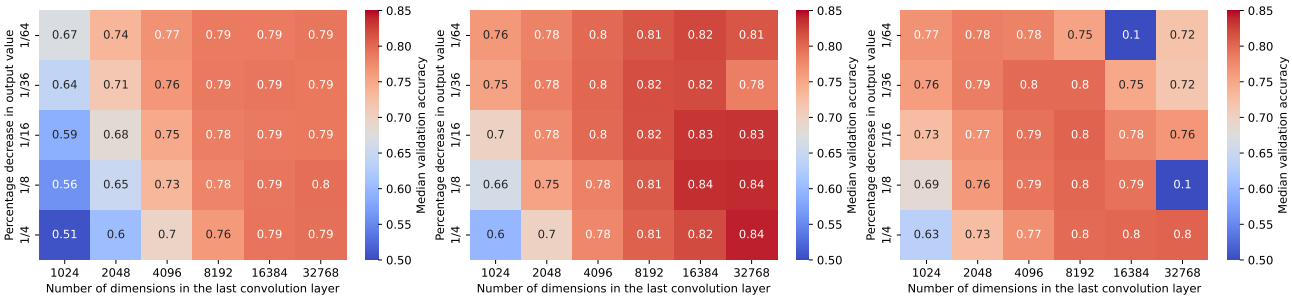
Figure 1: Decrease rate of CNN dimensionality

2. Simplified CNN and classification accuracy

In order to examine the accuracy of the zero-shot metric, we create a simple CNN to confirm the relationship with the classification accuracy.

The simple CNN consists of only a two-dimensional convolutional layers (Conv2D layer) and MaxPooling layers, GlobalAveragePooling layer (GAP layer) and Softmax layer. CNNs for classification tasks usually have a structure with dimensionally reduction from the first to the last layer. We consider CNNs with such a dimensionally-decreasing structure, solving classification tasks generally use a structure in which the number of dimensions decreases from the first intermediate layer to the output layer. Figure 1 show the number of dimensions in the first Conv2D layer is 65536 and the number of dimensions in the final Conv2D

ORCID iDs Chisato Takahashi: 0000-0002-0642-6550, Kenya Jin'no: 0000-0002-0431-5769



(a) 3 Conv2D layers (b) 6 Conv2D layers (c) 9 Conv2D layers

Figure 2: CNN Structure and classification accuracy rate

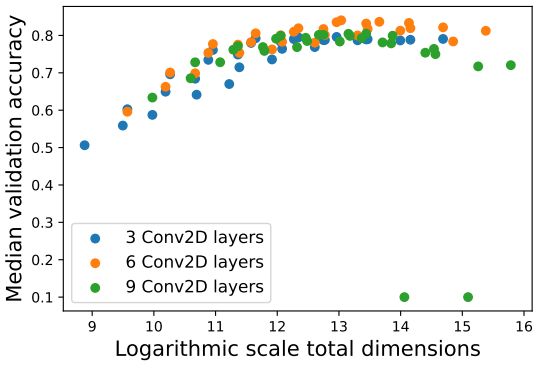


Figure 3: Total number of dimensions and classification accuracy rate

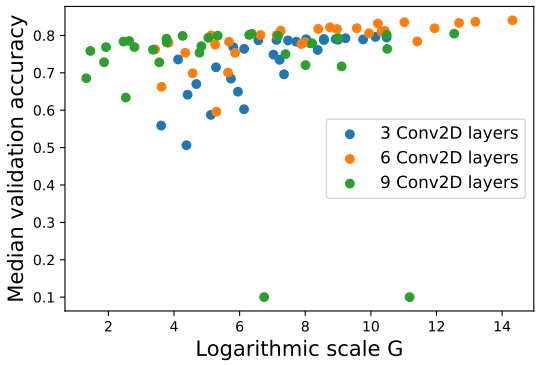


Figure 4: G value and classification accuracy rate

layer is 4096, so the reduction rate of the number of dimensions in the CNN is 1/16. We consider various reduction rate CNNs and the number of dimensions of the final Conv2D layer.

We experimentally examine the image classification capability of each CNN using CIFAR-10[6]. After 50 epochs with training with training data using CIFAR-10, the image classification accuracy of each CNN is calculated using test data. Nine trials are performed for each CNN structure, and the highest accuracy rate among the trials is applied.

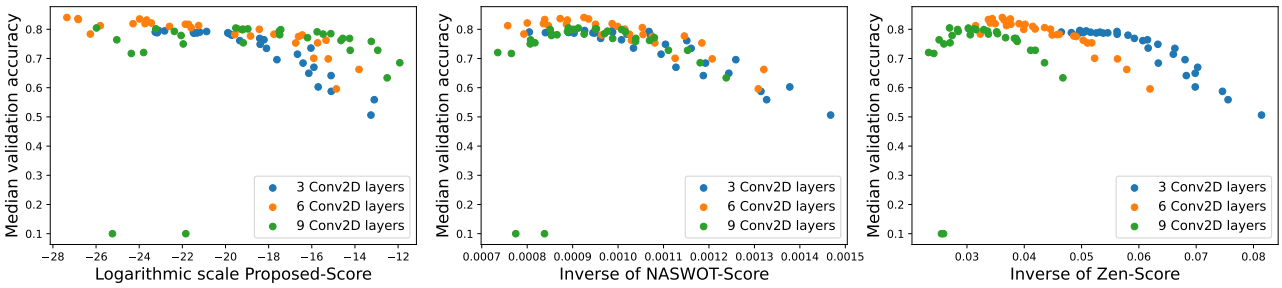
Figure 2 shows the relationship between the median accuracy of the CNN in the nine trials and the rate of increase in the number of dimensions of the CNN as the number of dimensions of the CNN when the dimensionality of the final Conv2D layer is varied. The vertical axis denotes the dimensionality reduction rate and the horizontal axis denotes the dimensionality of the final conv2D layer. The results as shown in Fig.2 indicate that the percentage of accuracy responses is higher when the dimensionality of the CNN is reduced at a higher rate.

3. Estimation of CNN classification accuracy

3.1. Total number of dimensions and classification accuracy

We examine whether it is possible to estimate the classification accuracy of a simple CNN based on the total number of CNN dimensions.

Figure 3 shows the relationship between the total number of CNN dimensions and the classification accuracy of the CNN. The horizontal axis denotes the total number of dimensions and the vertical axis denotes the median accuracy rate of the CNN in the nine trials. It can be seen that there is a positive correlation between the number of layer dimensions and the median accuracy rate, regardless of the number of layers in the Conv2D layer. However, the median accuracy rate drops when the total number of dimensions is very large, making it difficult to estimate a CNN with high classification accuracy based on the total number of dimensions alone.



(a) Proposed-Score (b) NASWOT-Score (c) Zen-Score

Figure 5: Evaluation Index Value and classification accuracy rate

3.2. Total GAP layer output and classification accuracy rate

We examine whether the accuracy of classification of a simple CNN can be estimated from the index value obtained by Eq. (1). Equation. (1) is the sum of the expected outputs of the GAP layer in the initial CNN, multiplied by the number of layers.

$$G = (\mathbb{E}_x \|f(x)\|_{L_1} + 1)^l, \tag{1}$$

where $f(x)$ is the output of the GAP layer when a standard Gaussian distribution x is input of CNN. $\mathbb{E}_x$ means the expected value for x, and l denotes the number of layers.

The relationship between the value of G and the classification accuracy of the CNN is shown in Fig.4. The horizontal axis is the value of G and the vertical axis is the median accuracy rate of CNN in the nine trials. There is a positive correlation between the value of $\mathbb{E}_x \|f(x)\|_{L_1}$ and classification accuracy. However, the value of $\mathbb{E}_x \|f(x)\|_{L_1}$ becomes smaller as the number of layers in the CNN increases. Therefore, we set $\mathbb{E}_x \|f(x)\|_{L_1} + 1 > 1$ and multiply this value to the power of l to suppress the effect of the number of layers. The larger the G value, the fewer CNNs have low classification accuracy. In other words, selecting a CNN with a larger G value than a certain value will reduce the possibility of having a CNN with low classification accuracy.

3.3. Proposal Evaluation Index

The evaluation index of CNN classification accuracy reflecting the total number of dimensions and G is shown in Eq. (2).

$$C = \frac{1}{d \cdot G}, \tag{2}$$

where d is the total number of Conv2D layer dimensions in CNN.

It is assumed that the closer the value of C is to 0, the better the classification accuracy of the CNN. We examine whether the classification accuracy of a simple CNN can be estimated from the value of C. The relationship between the value of C and the classification accuracy of CNNs is shown in Fig.5a. The horizontal axis is the value of C and the vertical axis is the median accuracy rate of the CNN in the nine trials. The smaller the value of C, the more accurate the CNN is, and it is considered possible to estimate the classification accuracy for a simple CNN.

4. Comparison

4.1. Accuracy of Evaluation

To compare with the proposed index, we derive the evaluation indicaions that have been proposed; NASWOT-Score[4] and Zen-Score[5].

NASWOT-Score is an index that evaluates the classification accuracy of a ReLU-based neural network based on the activity of its output when two types of input data are given.

Zen-Score is an index that evaluates the classification accuracy of a CNN using both the expected gradient of the output of the final Conv2D layer relative to its input and the per-channel variance statistics of the BatchNormalization layers (BN layers) after the CNN is transformed into a vanilla convolutional neural network (VCNN) with a BN layers added.

Fig.5b and Fig.5c show the relationship between the inverse of the evaluation index value in the previous study and the median accuracy response rate in the nine trials.

While the evaluation index values in the previous studies indicate that the larger the value is the better classification accuracy. On the other hand, our proposed method is the closer the value to 0 is the better the classification accuracy. In this experiment, to make the relationship between the values of the indexes easier to understand, the values of the indices in the previous studies are made inverse. Both of the evaluation indices in the previous studies show a negative correlation between the evaluation index value and the classification accuracy of the CNN. However, when the evaluation index value exceeds a certain value, the classification accuracy slightly decreases.

To identify which evaluation indexes can be used to find CNNs with high classification accuracy, we compare the

median and standard deviation of the classification accuracy rate for the CNNs in the top 10% of the index values. The results are shown in Table 1.

Table 1: CNN in the top 10% of index values

	Proposed Score	NASWOT Score	Zen Score
Median accuracy	80.5%	76.4%	75.0%
Standard deviation	0.223	0.213	0.274

Table 1 shows that the top 10% CNNs have the highest classification accuracy when the proposed evaluation index is used. The standard deviation is almost the same for all the evaluation indices. Therefore, in this experiment, the proposed evaluation index is the most suitable for finding a high-performing CNN.

4.2. Computation time

To reduce the NAS search time, the time used to evaluate the CNN should be as short as possible. We compare the time required to compute each metric value. Table2 shows the time required for the actual evaluation of CNNs.

Table 2: Computation time of evaluation index

Proposed-Score	NASWOT-Score	Zen-Score
1.96s	81.3s	2.39s

OS : Ubuntu 20.04.3 LTS
CPU : Intel(R) Xeon(R) W-2295 CPU @ 3.00GHz
GPU : NVIDIA RTX A6000 48GB
Computer memory : 384 GB

Table2 shows that the proposed evaluation index has the shortest computation time.

NASWOT-Score is considered to be very time-consuming because it requires replacing the output values of each layer with a binary vector in order to obtain the evaluation index value.

Zen-Score is considered to take a little longer than the proposed evaluation index because it requires both the product of the Frobenius norm of the difference between the outputs for two inputs and the variance statistics for each channel in the BatchNormalization layer(BN layer).

The proposed evaluation index requires three factors: the number of dimensions of the CNN, the number of layers, and the total output of the GAP layers. However, the computation time is expected to be short. This is because obtaining the number of dimensions and layers is very easy and the output of the GAP layer has fewer outputs than the Conv2D layer.

5. Conclusions

The experimental results suggest that a simple CNN consisting of only a Conv2D layer and a MaxPooling layer may be able to evaluate the classification accuracy of the CNN with the proposed evaluation index. In addition, it was found that the proposed evaluation metric can perform the evaluation in the shortest time among the metrics that can be evaluated without training.

In the future, we will examine how the proposed evaluation index captures the characteristics of simple CNNs. In addition, we will confirm whether the performance evaluation can be performed even for CNNs with skip structures, and aim to create a high-performance NAS.

Acknowledgments

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Number: 20K11978. Part of this work was carried out under the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

References

[1] Frank Hutter, Lars Kotthoff, Joaquin Vanschoren, AutoML: Methods, Systems, Challenges, Springer, 2019. https://www.automl.org/wp-content/uploads/2019/05/AutoML_Book.pdf

[2] Adanan Masood, Automated Machine Learning, Packt Publishing, 2021.

[3] Barret Zoph, Quoc V. Le, "Neural Architecture Search with Reinforcement Learning," Proc. ICLR 2017, https://arxiv.org/abs/1611.01578, 2016.

[4] Joseph Mellor, Jack Turner, Amos Storkey, Elliot J. Crowley, "Neural Architecture Search without Training," Proc. ICML 2021, https://arxiv.org/abs/2006.04647, 2020.

[5] Ming Lin, Pichao Wang, Zhenhong Sun, Hesen Chen, Xiuyu Sun, Qi Qian, Hao Li, Rong Jin, "Zen-NAS: A Zero-Shot NAS for High-Performance Deep Image Recognition," Proc. ICCV 2021, https://arxiv.org/abs/2102.01063, 2021.

[6] The CIFAR-10 dataset, http://www.cs.toronto.edu/ kriz/cifar.html