ZeroLess-DARTS: Improved Differentiable Architecture Search with Refined Search Operation and Early Stopping

Najmeh Fayyazifar
Harry Perkins Institute of Medical Research, University of Western Australia, Perth, Australia; School of Allied Health, Faculty of Health Sciences, Curtin University, WA, Australia

Selam Ahderom
School of Science, Edith Cowan University, Perth, Australia

Najmeh Samadiani
CSIRO Manufacturing, Private Bag 10, Clayton, Australia

Andrew Maiorana
School of Allied Health, Faculty of Health Sciences, Curtin University, WA, Australia

Girish Dwivedi
Harry Perkins Institute of Medical Research, University of Western Australia, Perth, Australia

ABSTRACT

Differentiable architecture search (DARTS) method has gained noticeable popularity in neural architecture search (NAS) domain as it reduces the required search time compared to reinforcement learning and evolutionary based NAS algorithms. However, some further studies have indicated that the search algorithm of DARTS may be suboptimal, and its performance may deteriorate over the search process. In this paper, we provided a detailed performance analysis of the DARTS search algorithm (on the CIFAR10 image classification task) from different aspects such as changes in accuracies of derived architectures at each search epoch, the trend of changes in strengths of different operations over successive epochs, and the number of skip connections per normal cells. We propose ZeroLess-DARTS that considerably improves original DARTS performance on the CIFAR10 dataset (Cohen's D = 2.213), by refining the operation space in the search procedure and introducing an early stopping criterion. We show that our approach is generalizable to time series classification tasks by evaluating the performance of our model on one-dimensional ECG signals for WCT (Wide Complex Tachycardia) classification (Cohen's D = 1.706).

CCS CONCEPTS

• **Artificial Intelligence, Machine Learning, Computer Vision**;

KEYWORDS

Neural architecture search, Differentiable architecture search, operation space, early stopping

ICMLC 2023, February 17–20, 2023, Zhuhai, China

ACM ISBN 978-1-4503-9841-1/23/02…$15.00
https://doi.org/10.1145/3587716.3587725

ACM Reference Format:
Najmeh Fayyazifar, Selam Ahderom, Najmeh Samadiani, Andrew Maiorana, and Girish Dwivedi. 2023. ZeroLess-DARTS: Improved Differentiable Architecture Search with Refined Search Operation and Early Stopping. In *2023 15th International Conference on Machine Learning and Computing (ICMLC 2023), February 17–20, 2023, Zhuhai, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3587716.3587725

1 INTRODUCTION

In the last decade, deep learning approaches and mainly Convolutional Neural Networks (CNN), became the dominant research paradigm for image classification [1], object detection [2], and machine translation [3] tasks. Moreover, CNN models perform well in disease identification tasks using an ECG recordings dataset [4-6]. CNNs have also been employed in other image classification tasks [7, 8]. CNNs can automatically extract features (by using filters) from unstructured data, and do not require manual feature engineering techniques that are commonly used in classical classifiers. However, the CNN architectures themselves are carefully hand-crafted to optimize feature extraction for a given task. The performance of these (hand-designed) CNN structures [9, 10] heavily relies on the prior knowledge of re-searchers in relation to CNN architecture design and optimization. In the last few years, the concept of Neural Architecture Search (NAS) has been increasingly studied to automatically design the best possible CNN model, for a given task, with the minimum human intervention [11]. Early research in NAS usually incorporated Reinforcement Learning (RL) [12] or evolutionary algorithms [13] to search for an optimized CNN architecture. These search methods were able to discover CNN models that provided state-of-the-art accuracy; however, they required huge computational resources. Differentiable architecture search methods were proposed to address the issue of the huge computation time required by RL and evolutionary methods to discover the best possible neural architecture. DARTS [14] transforms the discrete operations used in a convolutional neural network (used for image classification task on CIFAR10 dataset) to a continuous space by using a Softmax function (over all possible operations). Then, Stochastic Gradient Descent (SGD) is used to optimize over these continuous Softmax values.

DARTS achieved competitive results compared to RL-based and evolutionary-based methods, requiring only 4 GPU days on CIFAR10 search space to produce comparable results [14]. However, we discovered a few deficiencies in the DARTS algorithm. The search algorithm in DARTS shows a tendency to allocate higher Softmax values to the Zero operation, used to model the lack of connection between two nodes, (among all possible operations), and thereby the final architecture eventually becomes a disconnected graph (see Section 3.2 for more details).

The other deficiency of DARTS that has been identified in a few variants of the DARTS [15-19] (running DARTS unmodified) is the increase in the number of skip connections during the search procedure; however, to the best of our knowledge, there is no statistically conclusive evidence in relation to the number of skip connections that can provide the highest performance.

The overall goal of this paper is to analyze and understand the deficiencies of the DARTS algorithm and overcome them. To this end, we make the following contributions:

- We investigate and analyze the performance of the DARTS algorithm from a few different aspects and demonstrate the failure modes.
- Based on discovered deficiencies, we refine the search space of the DARTS algorithm.
- By statistically studying the DARTS algorithm, we provide support for the best number of skip connections.

The rest of this paper is organized as follows. In Section 2, we provide the basic concepts of the DARTS approach: concepts that we will refer to in the rest of this paper. Section 3 experimentally analyzes the DARTS method, including the evolution of the test accuracy, the relative strength of operations over the operation search space, and the number of skip connections over successive search epochs. Section 4 describes our proposed approach to improve the deficiencies discovered in the original DARTS method. Our results are discussed in Section 5 and conclusions are provided in Section 6.

2 BASIC CONCEPTS OF DIFFERENTIABLE ARCHITECTURE SEARCH (DARTS)

In this section, we provide an overview of the DARTS method [14]. This method proposes a CNN model for image classification task. In DARTS, the search space is defined as the allowed operations in their final models. Their CNN search space consists of two types of cells namely ‘Normal cell’ and ‘Reduction cell’, each cell containing standard operations (used in CNN). The final architecture is formed by stacking 20 cells. The reduction cell is designed to reduce the dimension of its input (by a factor of two) and is only used twice in the final architecture (at the first-third and the second-third of cells- cells number 7 and 14 out of 20 cells). Each of the cells is represented as a Directed Acyclic Graph (DAG) where the nodes of the graph are latent representations (i.e., feature maps in CNN), and edges are operations between two nodes. All cells have two input nodes (outputs from two previous cells), four intermediate nodes, and one output node that concatenates the outputs of intermediate nodes.

There are 8 allowed standard operations in the search space of the DARTS: separable convolution and dilated separable convolution (both with possible filter sizes of 3×3 and 5×5), max-pooling and average-pooling (both of size 3×3), skip connection (identity mapping), and the Zero operation (models the lack of connection between two nodes). These discrete operations are relaxed into a continuous space by computing the Softmax function over all possible operations. After operation relaxation (transforming discrete operations into continuous values), the task of architecture search is treated as jointly optimizing operations’ weight (Softmax values) and network’s weights, using SGD. In the DARTS paper, the search algorithm was performed for 50 epochs, and once the search was completed, the final architecture was pruned using Softmax values of operations in the last epoch.

3 DARTS PERFORMANCE ANALYSIS

3.1 Test Accuracy of Discovered Architectures Over Successive Search Epochs

In this section, we examine the finding in the Robust-DARTS paper [20], which identified some potential limitations of the original DARTS search procedure. Specifically, they found that the accuracy of the DARTS-derived architectures drops after epoch 40 of the search procedure. However, the authors of the Robust-DARTS performed the search strategy of the DARTS on a constrained search space (their defined S5 search space, Section III of [20]). They placed two sets of constraints: a) reduced the number of intermediate nodes (from 4 in the original DARTS method - described in Section 2) to only one node, b) constrained the number of allowed operations (from 8 operations in the original DARTS – Section 2) to only three operations (3×3 Sep-Conv, skip connection, and max pooling). To evaluate the search performance, they performed the search procedure, trained the derived architectures from scratch, and computed the test regret (defined as the difference between global minimum test error and reported test error for derived architecture at each epoch of the search algorithm). They observed that the test regret increases after epoch 40.

To further examine the finding of the authors of the Robust-DARTS [20], we repeated their experiments with the exact same settings of their method (search space and parameter settings). We trained from scratch all unique architectures that were derived during the execution of the search strategy (Fig.1). As illustrated in this figure, we reproduced their results where the test accuracy of unique architectures decreases after some number of epochs (after epoch 38). We should mention that over 50 epochs of search, using the limited search space of the Robust-DARTS, only 8 unique architectures were derived. For example, the architecture derived at epochs 2 and 3 were identical (shown as e2-3 in Fig.1). In cases where the architectures were identical for some epochs, we only trained that architecture once, and illustrated (Fig.1) the derived accuracy as the test accuracies of all those architectures. The authors of the Robust-DARTS [20] claimed that the observed increase in test regret (consequently drop in test accuracy) is due to the inefficiency of the DARTS search strategy. However, this claim needs further investigation to be validated. Specifically, more experiments are required to understand if the performance decay reported in [20] is really the impact of the search strategy of the DARTS algorithm,

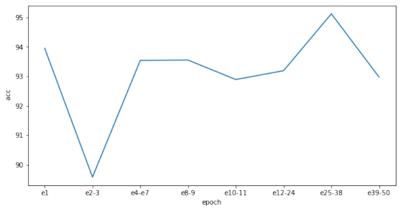


Figure 1: Reproduced test accuracies of derived architectures using the Robust-DARTS's constrained search space

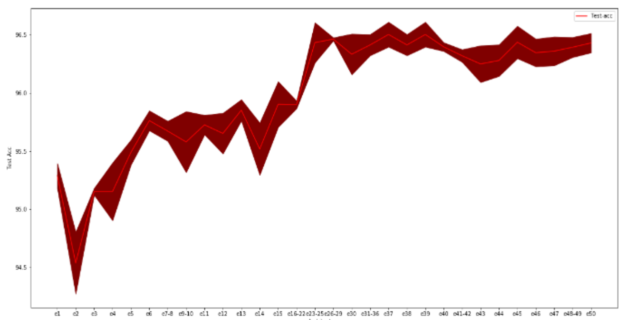


Figure 2: Reproduced test accuracies of derived architectures using DARTS's search space

or if the constrained search space is responsible for the observed accuracy drop.

To answer this question, we repeated the experiment with the original DARTS search space (larger than [20]) and settings, and we found that the drop in test accuracy disappeared (Fig.2 – fluctuating but almost stable after epoch 26). This figure shows that there is a definite but modest increase in the performance of the derived architectures over the search process. Since the performance of the DARTS doesn't drop over the search process (on the search space of the original DARTS), it appears that the performance decay observed in the Robust-DARTS paper [20] may be related to the constrained search space.

Overall, our experiments show that the performance of the DARTS search strategy (on their defined search space), doesn't drop over the original DARTS search period (50 epochs). However, as the final accuracy does not improve after a certain point (epoch 26 in our experiments), a properly designed early stopping can be used to terminate the search procedure. We thus propose an early stopping criterion, as introduced in section 4.1.

3.2 Operation Strength Over Successive Search Epochs

As described in section 2, in the DARTS algorithm, the task of architecture optimization is treated as the optimization of the Softmax values of operations (strength of operations) over the search epochs. Therefore, changes in the strength of operations, resulted from the SGD optimization algorithm, are the key factor in determining the final derived architecture. In this section, we empirically study these changes over search epochs. We performed the search procedure over the search space of the DARTS, and observed that the strength of the Zero operation becomes dominant over the search course, resulting in other operations' strengths being close to each other and moving toward zero (due to the Softmax function used for modeling architecture parameters – Section 2). Fig. 3. a shows the Softmax values between the last pair of nodes in Normal cells. This is one of the downsides of the DARTS algorithm. As DARTS selects the operation with the highest Softmax value as the operation between each pair of nodes, the high values of the Zero operation lead to a disconnected network (in our experiments, 12 out 14 possible operations were Zero operations after pruning).

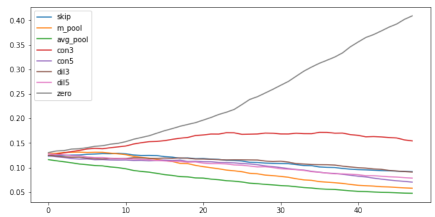
In this section, the impact of refining search operation (excluding Zero operation from search space) of DARTS is examined. More particularly, the search procedure of the DARTS algorithm is performed with the exact same settings as the DARTS original except excluding Zero operation from the search space. Fig.3. b shows the strength of other operations (between the same nodes as the previous experiment -last pair of nodes) over succeeding epochs.

As Fig.3. b shows, the final order of other operations (other than "Zero") is not identical to the experiment that included Zero operation (which leads to different derived architectures with different accuracies). Additionally, Fig. 3. b. shows that the Softmax values of the other operations haven't been compressed toward zero, thereby providing more freedom to compete and potentially end up with a better order (than the previous experiment). The derived architectures from both experiments (original DARTS and our refined search operation) and their test accuracies are shown in Fig. 4 and Section 5.1, respectively.

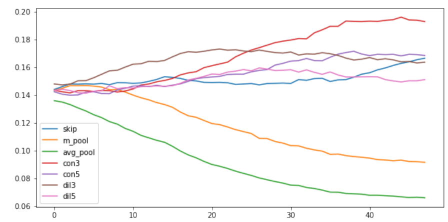
3.3 Number of Skip Connections Over Successive Search Epochs

A few recent studies [15-19] have shown that over the search procedure of the DARTS, the number of skip connections in the Normal cells increases significantly. To confirm this finding, we performed different runs of the search algorithm, and we show the results of three different runs in Fig. 5. This shows that the general trend of the number of skip connections is increasing and confirms the finding of [15-19]. However, in different runs of the search, the number of skip connections in the final derived architecture is different. These studies [15-19] variously claimed that the domination of skip connections could cause an accuracy drop (in discovered architectures) and suggested some modifications to the DARTS algorithm to mitigate this issue. However, to the best of our knowledge, there has not been a statistically conclusive investigation of the optimal number of skip connections that can provide the highest performance. For example, Progressive- DARST [15] and Fair- DARTS [16] limited the number of possible skip connections to a pre-defined value (M) and used trial and error to discover the best possible value

of M, using only three search runs (the trial and error suggested zero, one, and two as the best values of M). In section 4.1, we provide further insight (using a larger set of experiments) on the distribution of the number of skip connections in different search runs, and how this number impacts the accuracy of derived architectures.

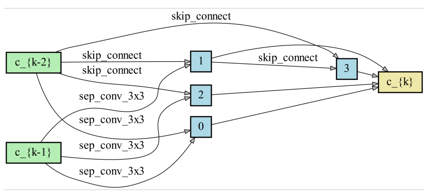


a) Softmax values of operations between the last pair of nodes including the “Zero” in the search procedure

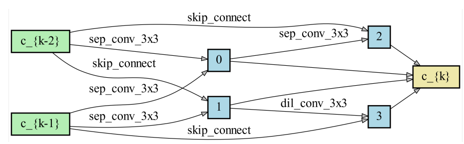


b) Softmax values of operations between the last pair of nodes excluding “Zero” in search procedure

Figure 3: Softmax values of operations last pairs of nodes

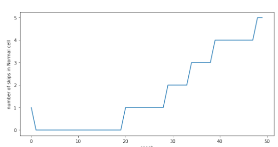


a) The derived normal cell architecture by including the “Zero” operation in the search space of the DARTS

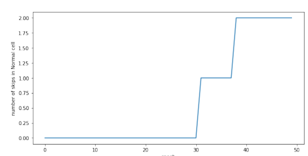


b) The derived normal cell architecture by excluding the “Zero” operation from the search space of the DARTS

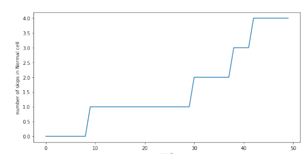
Figure 4: Final derived normal cell architectures



Run 1



Run 2



Run 3

Figure 5: Changes in the number of skip connections over successive search epochs of the DARTS algorithm for 3 different runs

4 OUR PROPOSED APPROACH

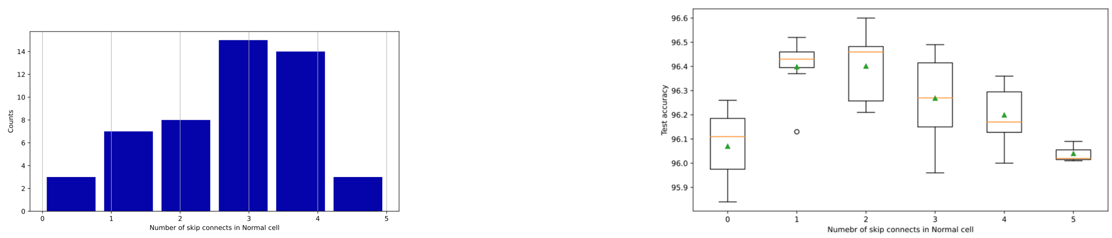
4.1 Discovering the Best Number of Skip Connections in Normal Cells

As mentioned in section 3.3, in different execution of the DARTS algorithm, different skip connections can appear in Normal cells (the focus is on Normal cells, as they mostly form the final architecture-18 out of 20 cells in the final architecture are Normal cells). In this section, through several experiments, a statistically conclusive support for the best number of skip connections per Normal cells is provided.

To this end, the DARTS search algorithm is executed 50 times with 50 different seeds. Similar to the DARTS procedure, for each execution, the architectures that were derived at the last epoch of the search (50th epoch as the original DARTS suggested) were selected, and the histogram of the number of skip connections in the derived Normal cell were computed. As can be seen in Fig. 6. a., most architectures have 3 or 4 skip connections in their Normal cell. To illustrate the impact of the number of skip connections (per Normal cell) on the test accuracy of these derived CNN architectures, each architecture is trained from scratch, and test accuracy is computed. The boxplot of accuracies, for each number of skip connections, are shown in Fig. 6. b. As the figure illustrates, the architectures with 1 (average accuracy of 96.39%) and 2 (average accuracy of 96.40%) skip connections reported the highest average accuracies (triangles on the graph). Other measures such as: median accuracy (orange line in the boxes), maximum accuracy (top line), third quartile accuracy (top line of the box), first quartile accuracy (bottom line of boxes), and minimum accuracy (bottom line) are higher for architectures with 1 and 2 skip connections compared to architectures with the other number of skip connections (zero, three, four and five).

To confirm these findings, in a second experiment, the relation between the number of skip connections and the average test accuracy (reported in Section 3. 1) is studied. As Fig. 7 shows, when the number of skip connections increases from zero (epoch 14) to two (epoch 26), test accuracy improves from 95.51% to 96.46%, and after epoch 26 no accuracy improvement is observed.

From the above two experiments, it can be concluded that one or two skip connections seem to optimally improve the performance, and increasing the skip connections beyond two (more particularly from three to five) does not seem to offer any advantage. Therefore, the search procedure should be continued until at least one skip connection presents in the Normal cell, and the search should be



a) Histogram of number of skip connections

b) Boxplot of test accuracy of architectures with the same number of skip connections

Figure 6: Histogram of number of skip connections in the normal cells produced by the DARTS over 50 different runs and Boxplot of test accuracy of architectures with the same number of skip connections

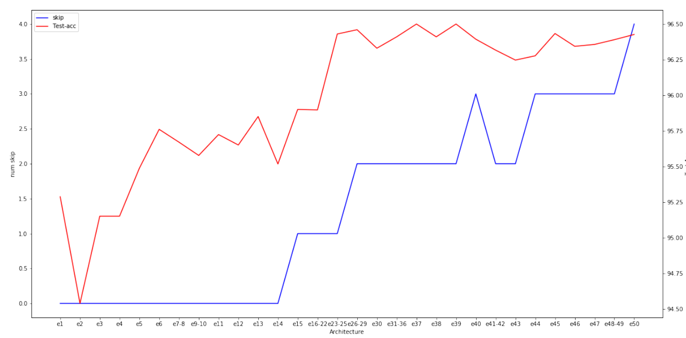


Figure 7: Relation between the accuracy of the derived architecture over the DARTS search procedure and the number of skip connections in their normal cell

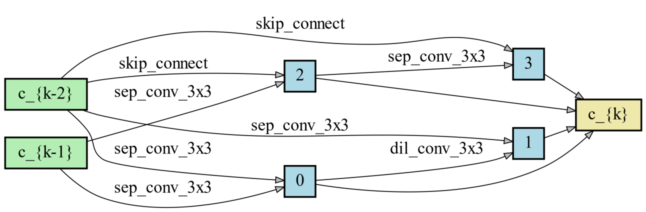


Figure 8: The architecture of our derived Normal cell, using our architecture selection criteria (on our refined search space)

terminated when the number of skip connections per Normal cells exceeds two.

4.2 ZeroLess-DARTS with Early Stopping

Based on our experiments and analysis, in this section, we proposed ZeroLess-DARTS with early stopping. As discussed in Section 3.2, Zero operation does not contribute to the original DARTS discretization procedure and when it is initially excluded from search space, the search algorithm discovers more accurate CNN architectures. To this end, the search space of our proposed ZeroLess-DARTS is refined to only contains 7 operations (Convolutions, Poolings and skip connection).

As suggested by our finding in Section 4.1, in our ZeroLess - DARTS with early stopping, the search procedure continues until at least one skip connection appears in each Normal cell and is terminated once the number of skip connections per Normal cell exceeds two. When the search is terminated, architectures with one and two skip connections are re-evaluated and the one with minimum validation loss is selected as the final CNN model. The derived architecture by ZeroLess-DARTS with early stopping is shown in Fig. 8

5 RESULTS AND DISCUSSION

5.1 Performance Analysis

In this section, firstly, the average test accuracy and standard deviation of our ZeroLess-DARTS model are compared against our reproduced DARTS results. Then, to confirm if our proposed search approach is robust (reproducible), the search results of our approach and the DARTS original algorithm are compared. Table 1 shows the average test accuracy and standard deviation of the proposed ZeroLessDARTS against our reproduced DARTS results.

As Table 1 illustrates, our ZeroLess-DARTS with early stopping outperforms the DARTS original algorithm, reaching from 96.36% to 96.64% accuracy. As the ZeroLess-DARTS derived improvements are modest, to measure the practical importance of our results, we computed the t-test and Cohen's D test. In inferential statistics,

Table 1: Accuracy, T-test and Cohen's D of our ZeroLess-DARTS and original DARTS reproduced results

Model	Accuracy ± SD (10 runs)	T-value P-value	Cohen'D test
DARTS model	96.36 ± 0.141	4.95 0.00010	D = 2.213
ZeroLess-DARTS	96.64 ± 0.09		

Table 2: Accuracy, T-test and Cohen's D of our ZeroLess-DARTS and the original DARTS reproduced results on search runs

Model	Accuracy ± SD (5 search runs)	T-value P-value	Cohen'D test
DARTS model	96.39 ± 0.132	3.88 0.0046	D = 2.258
ZeroLess-DARTS	96.71± 0.095		

Table 3: Comparison against the state of art methods- CIFAR10

Model	Test accuracy ± SD (3 runs)
DARTS	96.29 ± 0.17
Progressive-DARTS	96.45 ± 0.23
Robust-DARTS	96.34 ± 0.19
Fair-DARTS	96.28 ± 0.27
ZeroLess-DARTS with early stopping	96.66 ± 0.07

a key measure to determine whether the means of two sets are significantly different is a t- test [21]. Although a t-test verifies whether the difference between two data distributions is statistically significant, it doesn't give a measure of the 'effect size' of this difference (in practice how big this difference is). A commonly used method to compute the practical importance of an effect is Cohen's D test [22]. The t-test and the Cohen's D test results of ZeroLess-DARTS and original DARTS are summarized in Table 1.

The high t-value and very low p-value indicate that our results are statistically significant, and our ZeroLess-DARTS- derived model outperforms the DARTS-derived architecture considerably. The very large D value (as interpreted in [22, 23]) shows that the average accuracy of our ZeroLess-DARTS model has a very huge difference compared to the model derived by reproducing the DARTS original algorithm.

Table 2 illustrates the results of repeating the search procedure of ZeroLess-DARTS 5 times with 5 different seeds. The same procedure is performed for the original algorithm of the DARTS, and 5 different models are sampled. All models were trained from scratch only one time. This table confirms that over 5 different search runs, our ZeroLess-DARTS algorithm discovered more accurate models (compared to the DARTS original algorithm) and the standard deviation has been reduced. Moreover, Table 2 provides the t-test and the Cohen's D test results of ZeroLess-DARTS and original DARTS, computed over 5 executions of the search procedure. These results confirm that our proposed search approach is reproducible and can noticeably improve the DARTS method.

5.2 Comparison against state of the art

In this section, the performance of our proposed ZeroLess-DARTS is compared against state-of-the-art methods [14-16, 20]. We used their publicly available codes to reproduce their results. Comparison results on the CIFAR10 dataset are provided in Table 3.

5.3 Transferring to 1-dimensional ECG Recording Dataset

To validate the generalizability of ZeroLess-DARTS, we evaluated it on a dataset with a completely different nature, specifically, an ECG dataset containing WCT samples (collected at the University of Ottawa Heart Institute, Ontario, Canada). The dataset comprises 2906 Supraventricular Tachycardia (SVT) samples and 415 Ventricular Tachycardia (VT) recordings. We randomly selected 100 VT and 100 SVT samples for test and used the rest for training. We made some adjustments in the search operations of DARTS (replaced two-dimensional Convolution and Poolings with one-dimensional Convoluion and Poolings) to fit for one-dimensional ECG recordings. In Table 4, the average accuracy, standard deviation, t-test, and Cohen's D test results (for the models derived by ZeroLess-DARTS with early stopping and original DARTS) are provided.

As Table 4 shows, the model derived by ZeroLess-DARTS considerably outperforms DARTS derived model (t-value =2.36, p-value=0.04 and D= 1.706). This confirms that our proposed ZeroLess-DARTS is generalizable to other classification tasks, though with a completely different data nature (one-dimensional time series data).

6 CONCLUSION

In this paper, we proposed ZeroLess-DARTS (acc = 96.64 ± 0.09) with early stopping which significantly outperforms DARTS original algorithm (t-value = 4.95 and p-value = 0.00010 and Cohen's D = 2.213). To achieve this, we empirically studied the performance of

Table 4: T-test and Cohen's D test results of our ZeroLess-DARST and original DARTS reproduced results on WCT ECG recording dataset

Model	Accuracy ± SD (10 runs)	T-value P-value	Cohen'D test
DARTS model	87.61 ± 1.23	2.36 0.04	D = 1.706
ZeroLess-DARTS	89.14 ± 0.76		

the DARTS algorithm from different aspects such as: changes in accuracies of derived architectures at each search epoch, the trend of changes in strengths of different operations over successive search epochs, and the number of skip connections per Normal cells. We observed that the Zero operation that is used in the original DARTS does not contribute to architecture optimization.

Also, we discovered that the accuracy of derived architectures (over search epochs) does not deteriorate when performing the DARTS with the exact same settings (including 50 epochs of search run) suggested by the authors of the DARTS. However, through several experiments, we provided statistical support that architectures with one or two skip connections in their Normal cell provide the highest accuracy and after that no accuracy improvement is achieved.

Based on these findings, we refined the operation space of DARTS by excluding the Zero operation from the search space and performed the search procedure until at least one skip connection appeared in the Normal cells. The early stopping terminates the search algorithm once the number of skip connections per Normal cells exceeds two (the focus is on Normal cells, as they mostly form the final architecture- 18 out of 20 cells in final architecture are Normal cells). Then, architectures with one or two skip connections in their Normal cell are re-evaluated, and the one with minimum validation loss is selected as the final architecture. Some existing research considered constraining the number of skip connections to two, however, there was no guarantee that the architectures will contain at least one skip connection per Normal cell. Our experiments suggest that one skip connection is vital to discover an accurate CNN model. We evaluated the generalizability of our ZeroLess-DARTS with early stopping by transferring it into a dataset with a completely different nature (one-dimensional ECG recordings). Results confirm the generalizability of our approach (acc = 89.14 ± 0.76). Our ZeroLess-DARTS outperformed DARTS original algorithm (t-value= 2.36, p-value= 0.04, Cohen's D value=1.706).

REFERENCES

[1] Krizhevsky, A., Sutskever, I. and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25 (2012).

[2] Ren, S., He, K., Girshick, R. and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28 (2015).

[3] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q. and Macherey, K. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv: 1609.08144* (2016).

[4] Fayyazifar, N. An accurate CNN architecture for atrial fibrillation detection using neural architecture search. IEEE, City, 2021.

[5] Fayyazifar, N., Ahderom, S., Suter, D., Maiorana, A. and Dwivedi, G. *Impact of neural architecture design on cardiac abnormality classification using 12-lead ECG signals*. IEEE, City, 2020.

[6] Fayyazifar, N., Dwivedi, G., Suter, D., Ahderom, S., Maiorana, A., Clarkin, O., Balamane, S., Saha, N., King, B. and Green, M. S. A novel convolutional neural network structure for differential diagnosis of wide QRS complex tachycardia. *Biomedical Signal Processing and Control*, 81 (2023), 104506.

[7] Mrazova, I. and Kukacka, M. Image classification with growing neural networks. *International Journal of Computer Theory and Engineering*, 5, 3 (2013), 422.

[8] Thanh, T. T. P., Vununu, C., Atoev, S., Lee, S.-H. and Kwon, K.-R. Leukemia blood cell image classification using convolutional neural network. *International Journal of Computer Theory and Engineering*, 10, 2 (2018), 54-58.

[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. *Going deeper with convolutions*. City, 2015.

[10] He, K., Zhang, X., Ren, S. and Sun, J. *Deep residual learning for image recognition*. City, 2016.

[11] Bergstra, J., Yamins, D. and Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. PMLR, City, 2013.

[12] Bello, I., Zoph, B., Vasudevan, V. and Le, Q. V. *Neural optimizer search with reinforcement learning*. PMLR, City, 2017.

[13] Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y. L., Tan, J., Le, Q. V. and Kurakin, A. *Large-scale evolution of image classifiers*. PMLR, City, 2017.

[14] Hanxiao Liu, K. S., Yiming Yang DARTS: Differentiable architecture search. *International Conference on Learning Representations, ICLR 2019: New Orleans, Louisiana, United States* (2018).

[15] Chen, X., Xie, L., Wu, J. and Tian, Q. Progressive differentiable architecture search: Bridging the depth gap between search and evaluation. City, 2019.

[16] Chu, X., Zhou, T., Zhang, B. and Li, J. *Fair darts: Eliminating unfair advantages in differentiable architecture search*. Springer, City, 2020.

[17] Liang, H., Zhang, S., Sun, J., He, X., Huang, W., Zhuang, K. and Li, Z. Darts+: Improved differentiable architecture search with early stopping. *arXiv preprint arXiv:1909.06035* (2019).

[18] Yang, A., Esperança, P. M. and Carlucci, F. M. NAS evaluation is frustratingly hard. *arXiv preprint arXiv:1912.12522* (2019).

[19] Zhou, P., Xiong, C., Socher, R. and Hoi, S. C. H. Theory-inspired path-regularized differential network architecture search. *Advances in Neural Information Processing Systems*, 33 (2020), 8296-8307.

[20] Arber Zela, T. E., Saikia, T., Marrakchi, Y., Brox, T. and Hutter, F. *Understanding and robustifying differentiable architecture search*. City, 2020.

[21] Lowry, R. Concepts and applications of inferential statistics (2014).

[22] Cohen, J. *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[23] Sawilowsky, S. S. New effect size rules of thumb. *Journal of modern applied statistical methods*, 8, 2 (2009), 26.