Fitness Landscape Analysis of a Cell-based Neural Architecture Search Space

Devon Tao[1 [0009−0003−0507−0489]] and Lucas Bang[1 [0000−0003−2711−5548]]

Department of Computer Science, Harvey Mudd College, Claremont CA 91711, USA
{vtao, lbang} @hmc.edu

Abstract. Neural Architecture Search (NAS) research has historically faced issues of reproducibility and comparability of algorithms. To address these problems, researchers have created NAS benchmarks for NAS algorithm evaluation. However, NAS search spaces themselves are not yet well understood. To contribute to an understanding of NAS search spaces, we use the framework of *fitness landscape analysis* to analyze the topology search space of NATS-Bench, a popular cell-based NAS benchmark. We examine features of *density of states*, *local optima*, *fitness distance correlation* (FDC), *fitness distance rank correlations*, *basins of attraction*, *neutral networks*, and *autocorrelation* in order to characterize the difficulty and describe the shape of the NATS-Bench topology search space on CIFAR-10, CIFAR-100, and ImageNet16-120 image classification problems.

Keywords: Neural Architecture Search · fitness landscape · neural networks

1 Introduction

Neural networks have performed well in various tasks such as image classification [9, 13], speech recognition [1], and object detection [27]. However, achieving state-of-the-art performance has traditionally required expert knowledge of neural architecture design. This can pose a challenge for non-computer scientists who wish to use neural networks in their own work, but lack the specific neural network expertise, for example medical experts [24]. One recent solution to this is Neural Architecture Search (NAS), in which a neural architecture is algorithmically engineered as opposed to being hand-designed. NAS has shown to be an effective architecture design method, in some cases exceeding the performance of hand-designed architectures [36].

While NAS has achieved state-of-the-art performance, it has also faced reproducibility issues due to algorithmic complexity and expensive computational costs [14]. Furthermore, it has been difficult to compare different NAS methods due to differences in training procedures and search spaces [35]. To combat these problems, researchers have created NAS benchmarks, which provide common baselines for comparing algorithms and also significantly reduce the costs of

NAS evaluation [35, 7, 25]. One popular benchmark is NATS-Bench, a cell-based NAS search space [5].

Although there have been analyses of NAS search spaces as a whole [33, 3], there currently do not exist many deep analyses of particular NAS search spaces. While many NAS algorithms have performed well on these spaces [17, 2], there is a lack of understanding of the search spaces themselves. We aim to fill this gap by providing an analysis for the NATS-Bench topology search space through the framework of *fitness landscape analysis*, a concept that originates from biology [34] and has since been applied to optimization problems [18, 28]. We examine fitness landscape components of *density of states*, *local optima*, *fitness distance correlation* (FDC), *fitness distance rank correlations*, *basins of attraction*, *neutral networks*, and *autocorrelation* in order to characterize the difficulty and describe the shape of the NATS-Bench topology search space on three popular image classification datasets CIFAR-10, CIFAR-100 [12], and ImageNet16-120, which is a downsampled version of ImageNet [4]. We summarize our contributions as the following:

- We build upon previous analyses of NATS-Bench [20, 29] by providing a fitness landscape analysis of the NATS-Bench topology search space test accuracies.
- We calculate and analyze several components of the NATS-Bench topology fitness landscape, augmenting previous NATS-Bench analyses of density of states, FDC, and local optima networks [20] with additional characteristics of fitness distance rank correlations, basins of attraction, neutral networks, and autocorrelation. To our best knowledge, we are the first to calculate these metrics for the NATS-Bench topology search space.

2 Related Work

The first NAS method used reinforcement learning [36]. Since then, researchers have developed a variety of approaches, such as neuroevolution [26], differentiable architecture search [15], one-shot NAS [6, 8], and training-free methods [17, 2].

As for the search spaces themselves, there have been a number of established benchmarks for image classification problems such as NAS-Bench-101 [35], NAS-Bench-201 [7], NAS-Bench-301 [25], and NATS-Bench [5]. More recently, there have also been NAS benchmarks in other areas such as automated speech recognition [16] or natural language processing [11].

The research area of NAS search space analysis is still in its early stages, so there are currently few deep analyses of these spaces. The authors of NAS-Bench-101 perform a preliminary fitness landscape analysis of their benchmark, including aspects such as FDC, locality, and autocorrelation [35]. Traoré et al. expand on this analysis by introducing the concept of a fitness landscape footprint that additionally introduces the aspects of ruggedness, cardinality of optima, and persistence [30]. There have been a few analyses examining the NATS-Bench benchmark specifically. Thomson et al. examine the local optima networks of the size search space of NATS-Bench [29] and Ochoa and Veerapen perform a

fitness landscape analysis on the validation accuracies of the NATS-Bench topology search space. Their analysis examines aspects of density of states, FDC, and local optima networks [20]. We expand on the work of Ochoa and Veerapen by analyzing the test accuracies of the NATS-Bench topology search space, and by providing additional analyses of fitness distance rank correlations, basins of attraction, neutral networks, and autocorrelation.

3 Background

In this section, we introduce NATS-Bench, a repeated cell-based neural architecture search space. We then define a fitness landscape and its components and describe the specific fitness landscapes for the NATS-Bench topology search space.

3.1 NATS-Bench

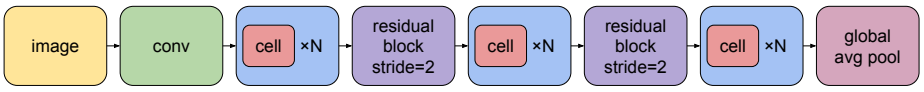

Fig. 1. Macro structure of a neural architecture in the topology search space of NATS-Bench. Visual based on original paper [5].

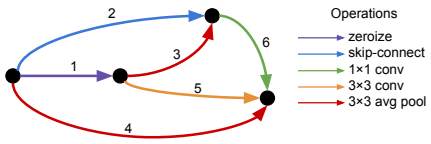

Fig. 2. DAG representation of an individual cell. Visual based on original paper [5].

NATS-Bench is a repeated-cell neural architecture benchmark that consists of two search spaces, a size search space $\mathcal{S}_s$ and a topology search space $\mathcal{S}_t$ [5]. We examine the topology search space $\mathcal{S}_t$ for our fitness landscape analysis, which is the same as NAS-Bench-201 [7]. The macro structure of each neural architecture begins with a 3-by-3 convolution with 16 output channels and a batch normalization layer. It is then followed by three stacks of $N = 5$ cells with a residual block between each cell stack. The number of output channels are 16, 32, and 64 for the three stacks respectively. These cell stacks are then followed by a global average pooling layer. The NATS-Bench architectures are trained

on CIFAR-10, CIFAR-100 [12], and ImageNet16-120, which is a downsampled version of ImageNet [4]. Performance data for architectures in $\mathcal{S}_t$ trained on 12 or 200 epochs of data can be accessed by querying the NATS-Bench API.[1] Additionally, architectures undergo many trials. For our landscape analysis, our fitness values are the test accuracies of architectures trained on 200 epochs of data, averaging over all trials.

Each cell in $\mathcal{S}_t$ can be represented as a densely-connected DAG with four vertices, where there is an edge from the ith node to the jth node if $i < j$ for a total of six edges. Each edge is selected from one of five operations: zeroize, skip connection, 1-by-1 convolution, 3-by-3 convolution, and 3-by-3 average pooling layer, where the zeroize operation represents dropping the edge. Then, there are $5^6 = 15625$ total architectures. However, some architectures are isomorphic, so there are only 6466 unique architectures [7].

3.2 Fitness Landscape Analysis

Definition We use the definition of fitness landscape provided by Pitzer et al. [21]. There is a solution space $\mathbf{S}$ and an encoding of the solution space $\mathscr{S}$. There is also a fitness function $f : \mathscr{S} \to \mathbb{R}$ that assigns a real-valued number to a solution candidate, and a distance metric $d : \mathscr{S} \times \mathscr{S} \to \mathbb{R}$. Then, a fitness landscape is defined as the tuple:

$$\mathscr{F} = (\mathscr{S}, f, d) \tag{1}$$

Fitness Landscape of NATS-Bench Topology Search Space Each architecture in the NATS-Bench topology search space can be represented as a string of length six where each character represents an edge operation for a corresponding edge in the DAG representation of the cell. Then for $\mathcal{S}_t$, $\mathscr{S}$ is the set of all possible string representations of a neural architecture. The distance function d is the Hamming distance between two such strings. We define the *neighborhood* of a solution candidate as $\mathcal{N}(x) = \{y \in \mathscr{S} | d(x, y) = 1\}$, that is, the set of architecture strings that represent a change of one edge operation from the architecture of x. We have three fitness functions, corresponding to average test accuracies of architectures trained on 200 epochs of data on CIFAR-10, CIFAR-100, and ImageNet16-120. Thus, we have three fitness landscapes, one for each image classification dataset.

For the purposes of analysis on NATS-Bench, we deviate slightly from Pitzer et al.'s definitions of phenotype and genotype. Because some architectures are isomorphic, in addition to the string representation of the architecture, each architecture also has a string representation of the unique isomorph. We consider the string representation of the architecture the *genotype*, and the string representation of the unique isomorph the *phenotype*. Due to numerical error, two architectures with the same phenotype may have two different fitnesses [7].

[1] https://github.com/D-X-Y/NATS-Bench

Density of States A density of states analysis examines the number of solution candidates with a certain fitness value. The density of states can tell us how likely it is to find a "good" solution via random search [23]. For example, a fitness landscape with many fitnesses close to the global optimum will be easier for random search than a fitness landscape with few fitnesses close to the global optimum.

Fitness Distance Correlations One measure of problem difficulty is the correlation between distances to the nearest global optimum (in our case, maximum) and the fitnesses of solution candidates. This correlation can help us measure the extent to which there is a "gradient" of fitness to a global optimum. One established metric is fitness distance correlation, which is a measure of problem difficulty for genetic algorithms [10]. If we let F represent a list of fitnesses of $\mathscr{S}$ and D represent the corresponding distances to the nearest global optimum, then the FDC is simply the Pearson correlation coefficient between F and D:

$$\text{FDC} = \frac{\text{cov}(F, D)}{\sigma_F \sigma_D} \tag{2}$$

Where $\text{cov}(F, D)$ is the covariance of F and D, and σ_F and σ_D are the standard deviations of F and D respectively. In addition to FDC, we also examine Spearman rank correlation and Kendall rank correlation between F and D. The Spearman fitness distance rank correlation is:

$$\rho = \frac{\text{cov}(R(F), R(D))}{\sigma_{R(F)} \sigma_{R(D)}} \tag{3}$$

where $R(F)$ and $R(D)$ are F and D converted to ranks, respectively. Then, the Kendall fitness distance rank correlation is:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} \text{sgn}(f_i - f_j)\text{sgn}(d_i - d_j) \tag{4}$$

where $n = |\mathscr{S}| = |F| = |D|$ and f_i and d_i are the ith elements of F and D, respectively.

We clarify that while the Spearman and Kendall correlations are correlations between fitness and distance, the term "fitness distance correlation" or FDC specifically refers to Pearson's correlation, as established in the literature [10].

Local Optima A solution candidate x is a local optimum if it is the fittest among its neighborhood [21].

$$\text{local optima}(x) \iff \forall y \in \mathcal{N}(x), f(x) > f(y) \tag{5}$$

The number of local optima can tell us about the global ruggedness of a fitness landscape, for instance, a multi-modal landscape is more globally rugged than a unimodal one. Furthermore, examining correlations between local optima fitness and distance to a global optimum can tell us the extent to which there is a progression of fitness from local optima to a global optimum.

Basins of Attraction Related to local optima is the concept of basins of attraction. Although the fitness landscape of NATS-Bench is a maximization problem, we still use the term *basins of attraction* in order to remain consistent with the literature. To understand basins of attraction, we must first understand an *upward path* to a local maximum. We adapt this definition for our maximization problem from the definition of a *downward path* by Pitzer et al. [22]. An upward path $p_\uparrow$ from candidate x_0 to x_n is the sequence $\{x_i\}_{i=0}^{n}$ where $(\forall i < j)$, $f(x_i) \leq f(x_j)$, $f(x_0) < f(x_n)$, and $x_{i+1} \in \mathcal{N}(x_i)$, that is, each solution candidate in the upward path is at least as fit as the previous one. Then, the weak basin of a local optimum o is defined as:

$$b(o) := \{x | x \in \mathscr{S}, p_\uparrow(x, o)\} \tag{6}$$

which is the subset of the search space that has an upward path leading to o. A strong basin of a local optimum o is defined as:

$$\hat{b}(o) := \{x | x \in b(o), (\nexists o' \neq o \in \mathscr{O}) \text{ they have } x \in b(o')\} \tag{7}$$

where $\mathscr{O}$ is the set of all local optima. In other words, the strong basin of a local optimum o is the subset of the search space that has an upward path only to o.

The relative fitnesses of the local optima combined with the relative sizes of their basins of attraction can give some indication of problem difficulty, as local optima with larger basins are more likely to be found via local search methods.

Neutral Networks A neutral network is a set of connected solution candidates with equal fitness and can be intuitively described as a "plateau" of fitness. The *percolation index* of a neutral network is the number of unique fitness values surrounding the neutral area [21].

Autocorrelation and Correlation Length Autocorrelation and correlation length are two measures for ruggedness of a fitness landscape [32]. The autocorrelation function for some lag i is the Pearson correlation coefficient between a random walk on the landscape and the same walk with time delay i. Then for a random walk F_t and a lag i, the autocorrelation function is:

$$\rho(i) = \frac{\text{cov}(F_t, F_{t+i})}{\sigma_{F_t} \sigma_{F_{t+i}}} \tag{8}$$

where F_{t+i} is F_t with a lag of i, $\text{cov}(F_t, F_{t+i})$ is the covariance of F_t and F_{t+i}, and σ_{F_t} and $\sigma_{F_{t+i}}$ are the standard deviations of F_t and F_{t+i}, respectively. Correlation length is defined as $\tau = \frac{-1}{\ln |\rho(1)|}$ for $\rho(1) \neq 0$, which is the expected distance between points before they become "uncorrelated" [32, 28].

4 Results

We compare the difficulty and shape of the NATS-Bench topology search space for three different fitness landscapes of CIFAR-10, CIFAR-100, and ImageNet16-120 test accuracies. We calculate, analyze, and visualize characteristics of density

of states, FDC and fitness distance rank correlations, local optima, basins of attraction, neutral networks, and autocorrelation. This analysis reveals that the use of different metrics results in different orderings of the difficulties of the three fitness landscapes.

While landscape analyses of larger search spaces may need to use sampling in order to approximate characteristics of the fitness landscape [19, 30], $\mathcal{S}_t$ in NATS-Bench is relatively small, so we are able to exhaustively evaluate the search space for most of our metrics. To estimate autocorrelation, we average 200 random walks of length 100 on the search space with random starting points. Our data and code are publicly available online.[2]

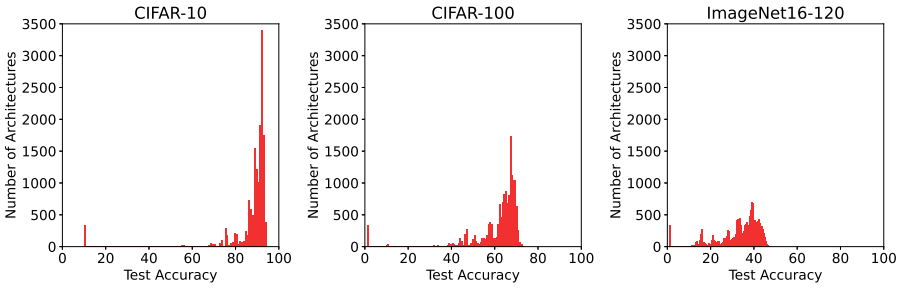

Fig. 3. Density of states of NATS-Bench topology search space test accuracies. The maximum fitness for each fitness landscape is 94.37, 73.51, and 47.31 for CIFAR-10, CIFAR-100, and ImageNet16-120 respectively.

4.1 Density of States

From Figure 3, we can see that CIFAR-10 has the most architectures near the global optimum, followed by CIFAR-100, then ImageNet16-120. This may indicate that NAS on architectures for CIFAR-10 image classification is the easiest, followed by CIFAR-100 and lastly ImageNet16-120. This order of difficulty for the NATS-Bench fitness landscapes matches the order of difficulty for the image classification problems themselves.

4.2 Fitness Distance Correlations

FDC can be used to characterize problem difficulty for genetic algorithms and can divide problems into three broad categories. FDC ≥ 0.15 is considered *misleading*, because solution candidates decrease in fitness as they approach a global optimum. $-0.15 <$ FDC < 0.15 is *difficult* because there is weak to no correlation between fitnesses and distances to the global optimum, and FDC ≤ -0.15

[2] https://github.com/v-tao/nats-bench-landscape

Table 1. Correlations between architecture fitness and distance to the global optimum.

	CIFAR-10	CIFAR-100	ImageNet16-120
FDC	-.2199	-.3090	-.3163
Spearman's ρ	-.4144	-.4666	-.3270
Kendall's τ	-.3200	-.3630	-.2502

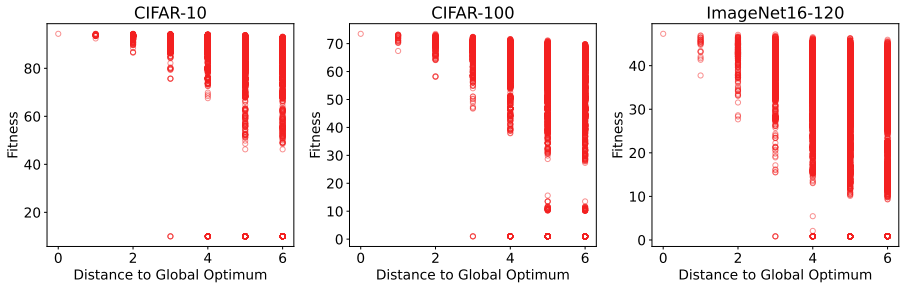

Fig. 4. Fitness vs. distance to the global optimum.

is *straightforward*, as solution candidates approaching the global optimum also increase in fitness [10]. Examining FDC, it would appear that ImageNet16-120 is the most straightforward landscape for genetic algorithms, followed by CIFAR-100 and then CIFAR-10. However, if we examine the rank correlations, it would appear that CIFAR-100 is the most straightforward, followed by CIFAR-10 and then ImageNet16-120.

4.3 Local Optima

Table 2. Correlations between architecture fitness and distance to global optimum of local optima.

	CIFAR-10	CIFAR-100	ImageNet16-120
# local maxima	17	24	36
FDC	-.8741	-.8225	-.6172
Spearman's ρ	-.8650	-.7916	-.6311
Kendall's τ	-.7223	-.6845	-.5050

Compared to the whole space of architectures, which has a weak negative correlation between fitness and distance to the global optimum, the subset of just the local optima displays a strong negative correlation between these two features. This suggests that there is generally a progression of fitnesses from local optima to the global optimum for all three fitness landscapes.

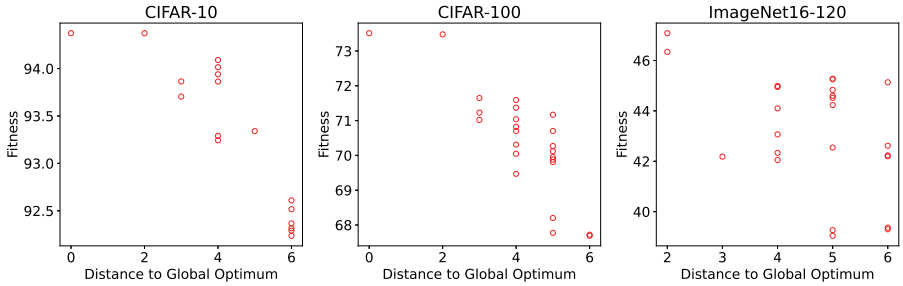

Fig. 5. Fitness vs. distance to the global optimum for local optima.

Table 3. Summary statistics of basins of attraction.

	CIFAR-10	CIFAR-100	ImageNet16-120
# weak basins	17	24	36
Average weak basin size	13122.06	13245.08	14337.81
Weak basin extent	.9989	.9984	.9977
# strong basins	4	8	14
Average strong basin size	10.00	7.88	2.29
Strong basin extent	.0026	.0040	.0020

4.4 Basins of Attraction

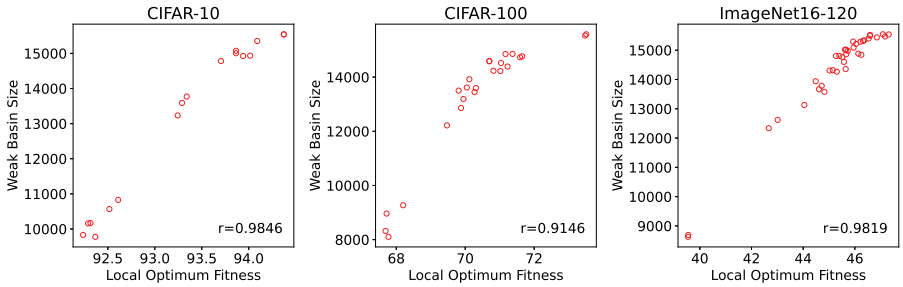

Fig. 6. Fitness vs. size of weak basin for local maxima.

For each of the fitness landscapes, the vast majority of the search space is in a weak basin of attraction, meaning that local optima can be reached via local search methods starting from almost any architecture. Combined with Fig. 5, this demonstrates that search on $\mathcal{S}_t$ for CIFAR-10 is easy, as the local optima for CIFAR-10 are close in fitness to the global optimum. While the ranges in local optima fitness are greater for CIFAR-100 and ImageNet16-120, all three

fitness landscapes show a strong correlation between local optima fitness and weak basin extent. This may be another indication of problem easiness, as fitter optima are more likely to be achieved via local search.

4.5 Neutral Networks

Table 4. Neutral network summary statistics.

	CIFAR-10	CIFAR-100	ImageNet16-120
# neutral networks	249	35	46
Average neutral network size	3.41	7.46	5.41
Maximum neutral network size	341	63	67

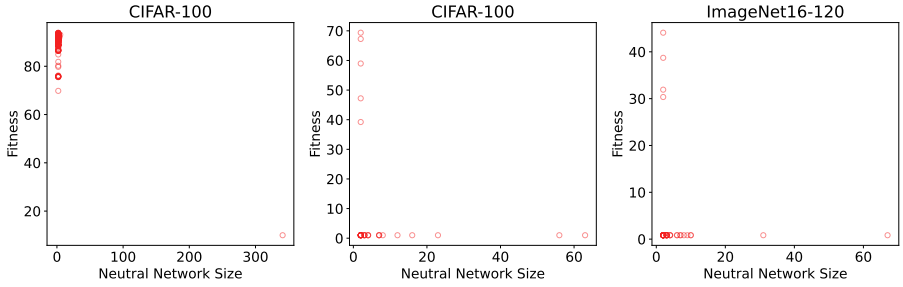

Fig. 7. Fitness vs. size of neutral networks.

A small fraction ($< .01$) of the search space belongs to a neutral network (a set of connected solution candidates with equal fitness), and furthermore, the majority of neutral networks are small, containing only a handful of architectures. Additionally, most architectures in neutral networks are the worst-performing architectures. These architectures are significantly worse than the rest, corresponding with the "spike" on the left of each histogram in Fig. 3. Upon closer inspection of the largest neutral network for each of the fitness landscapes, we see that these neutral networks consist entirely of architectures where the input node and output node are disconnected, resulting in an architecture that performs equal to random choice. In some circumstances, large neutral networks may be beneficial because they allow exploration of a large space while maintaining the same fitness [22, 31]. Although our data in Table 5 indicate that exploring these neutral networks do provide access to many genetic variations, these neutral networks consist of the worst architectures in the search space, so it may not be desirable to remain in these neutral networks over many iterations.

Table 5. Properties of largest neutral network and its neighbors.

	CIFAR-10	CIFAR-100	ImageNet16-120
Size	341	63	67
Fitness	10.00	1.00	.83
Max edit distance	6	6	6
Average edit distance	4.8031	4.8249	4.8318
Unique phenotypes	5	5	5
Unique neighbor genotypes	2868	917	974
Unique neighbor phenotypes	513	177	182
Percolation index	1707	690	745

4.6 Autocorrelation

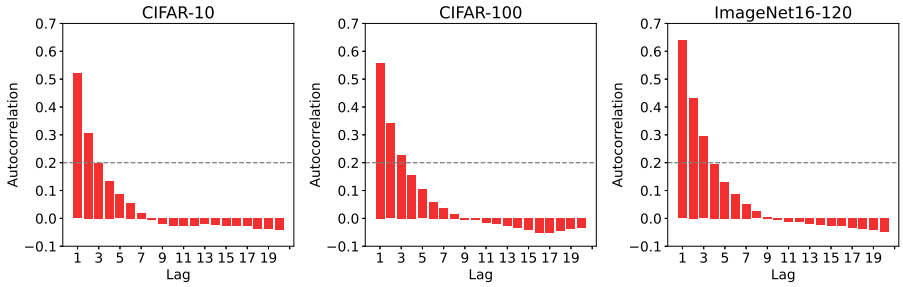

Fig. 8. Autocorrelation functions sampled from 200 random walks of length 100.

The correlation lengths are 1.53, 1.71, and 2.23 for CIFAR-10, CIFAR-100, and ImageNet16-120, respectively. Furthermore, we can see that the autocorrelation function for ImageNet16-120 decays slower than for CIFAR-10 or CIFAR-100. This indicates that at the local level, ImageNet16-120 is the smoothest out of the three fitness landscapes.

5 Discussion

Our fitness landscape analyses could indicate that the difficulties associated with the three fitness landscapes of CIFAR-10, CIFAR-100, and ImageNet16-120 on the topology search space of NATS-Bench correspond to the difficulties of the image classification problems themselves. This is reflected in the density of states, as CIFAR-10 has the greatest proportion of architectures close to the global optimum, followed by CIFAR-100 and ImageNet16-120. In addition, while all three fitness landscapes have similar weak and strong basin extents, both the number of local optima and the range of fitness values for local optima is smallest for

CIFAR-10, then CIFAR-100, then ImageNet16-120. This means that for CIFAR-10, not only is the global optimum more likely to be reached via local search, but also any local optimum reached is closer in fitness to the global optimum than for the other two fitness landscapes. Previous work indicates that ImageNet16-120 is the most difficult for validation accuracies of $\mathcal{S}_t$ [20], and our work shows a similar case for the test accuracies.

However, there are some discrepencies in our data which would at first appear to be contradictions. The progression of difficulty from CIFAR-10 to ImageNet16-120 is not supported by our data on correlations between architecture fitness and distance to the global optimum. By FDC, which is Pearson's correlation, ImageNet16-120 would be considered the most straightforward, while examining the rank correlations would point to CIFAR-100 as the most straightforward. We can resolve this discrepency by considering the different algorithms that may be applied to the optimization problem. The lowest FDC may indicate that ImageNet16-120 is the most straightforward for a genetic algorithm, whereas the lowest rank correlation may indicate that CIFAR-100 is the most straightforward for algorithms that only care about relative fitness values, for example hill climbing.

Additionally, previous work has used modality to describe the ruggedness of the NATS-Bench topology fitness landscapes [20]. While our own analysis of local optima support these claims of ruggedness, we should be more careful in describing this as *global* ruggedness. As ImageNet16-120 has the most local optima, followed by CIFAR-100 and CIFAR-10, it would appear that ImageNet16-120 is the most rugged on a global level. However, our autocorrelation and correlation length analyses point to the reverse order of ruggedness on a *local* level. These discrepancies demonstrate that the fitness landscape of any particular problem is multi-faceted, and many metrics are required to paint a fuller picture of a fitness landscape. Furthermore, a fitness landscape analysis should be done with nuance, and consider the different implications of metrics for different algorithms.

6 Conclusion

We performed a fitness landscape analysis of the topology search space of NATS-Bench [5], analyzing and visualizing features of density of states, FDC and fitness distance rank correlations, local optima, basins of attraction, neutral networks, and autocorrelation. Our analyses indicated that the problem difficulty of search on the topology search space of NATS-Bench for architectures that can perform well on CIFAR-10, CIFAR-100, and ImageNet16-120 datasets may correspond to the difficulties of the image classification problems themselves. We also demonstrated the importance of multiple metrics and nuance in the interpretation of a fitness landscape.

While these metrics can help to characterize the fitness landscape, ultimately they are not exact. Future work may include the comparison of different algorithms on NATS-Bench in order to discern how useful these metrics are for describing the true fitness landscape of NATS-Bench. As our current understanding

of NAS search spaces is limited, future work may also include fitness landscape analyses of other NAS search spaces, such as non-tabular search spaces [25] or for problems other than image classification [11, 16]. Another possible direction for future work is to investigate what properties of the architectures themselves cause the fitness landscapes to appear this way.

Acknowledgments. The authors would like to thank George Montañez for feedback on an earlier draft of this paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdel-Hamid, O., Mohamed, A.r., Jiang, H., Deng, L., Penn, G., Yu, D.: Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing **22**(10), 1533–1545 (2014). https://doi.org/10.1109/TASLP.2014.2339736
2. Chen, W., Gong, X., Wang, Z.: Neural architecture search on imagenet in four gpu hours: A theoretically inspired perspective. arXiv preprint arXiv:2102.11535 (2021). https://doi.org/10.48550/arXiv.2102.11535
3. Chitty-Venkata, K.T., Emani, M., Vishwanath, V., Somani, A.K.: Neural architecture search benchmarks: Insights and survey. IEEE Access **11**, 25217–25236 (2023). https://doi.org/10.1109/ACCESS.2023.3253818
4. Chrabaszcz, P., Loshchilov, I., Hutter, F.: A downsampled variant of imagenet as an alternative to the cifar datasets. arXiv preprint arXiv:1707.08819 (2017). https://doi.org/10.48550/arXiv.1707.08819
5. Dong, X., Liu, L., Musial, K., Gabrys, B.: Nats-bench: Benchmarking nas algorithms for architecture topology and size. IEEE transactions on pattern analysis and machine intelligence **44**(7), 3634–3646 (2021). https://doi.org/10.1109/TPAMI.2021.3054824
6. Dong, X., Yang, Y.: One-shot neural architecture search via self-evaluated template network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3681–3690 (2019). https://doi.org/10.1109/ICCV.2019.00378
7. Dong, X., Yang, Y.: Nas-bench-201: Extending the scope of reproducible neural architecture search. arXiv preprint arXiv:2001.00326 (2020)
8. Guo, Z., Zhang, X., Mu, H., Heng, W., Liu, Z., Wei, Y., Sun, J.: Single path one-shot neural architecture search with uniform sampling. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16. pp. 544–560. Springer (2020). https://doi.org/10.1007/978-3-030-58517-4_32
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
10. Jones, T., Forrest, S., et al.: Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In: ICGA. vol. 95, pp. 184–192 (1995)
11. Klyuchnikov, N., Trofimov, I., Artemova, E., Salnikov, M., Fedorov, M., Filippov, A., Burnaev, E.: Nas-bench-nlp: neural architecture search benchmark for natural language processing. IEEE Access **10**, 45736–45747 (2022). https://doi.org/10.1109/ACCESS.2022.3169897

12. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Tech. Rep. 0, University of Toronto, Toronto, Ontario (2009), https://www.cs.toronto.edu/ kriz/learning-features-2009-TR.pdf
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)
14. Li, L., Talwalkar, A.: Random search and reproducibility for neural architecture search. In: Uncertainty in artificial intelligence. pp. 367–377. PMLR (2020)
15. Liu, H., Simonyan, K., Yang, Y.: Darts: Differentiable architecture search. arXiv preprint arXiv:1806.09055 (2018). https://doi.org/10.48550/arXiv.1806.09055
16. Mehrotra, A., Ramos, A.G.C., Bhattacharya, S., Dudziak, Ł., Vipperla, R., Chau, T., Abdelfattah, M.S., Ishtiaq, S., Lane, N.D.: Nas-bench-asr: Reproducible neural architecture search for speech recognition. In: International Conference on Learning Representations (2020)
17. Mellor, J., Turner, J., Storkey, A., Crowley, E.J.: Neural architecture search without training. In: International Conference on Machine Learning. pp. 7588–7598. PMLR (2021)
18. Merz, P., Freisleben, B.: Fitness landscape analysis and memetic algorithms for the quadratic assignment problem. IEEE Transactions on Evolutionary Computation **4**(4), 337–352 (2000). https://doi.org/10.1109/4235.887234
19. Nunes, M., Fraga, P.M., Pappa, G.L.: Fitness landscape analysis of graph neural network architecture search spaces. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 876–884 (2021). https://doi.org/10.1145/3449639.3459318
20. Ochoa, G., Veerapen, N.: Neural architecture search: a visual analysis. In: International Conference on Parallel Problem Solving from Nature. pp. 603–615. Springer (2022). https://doi.org/10.1007/978-3-031-14714-2_42
21. Pitzer, E., Affenzeller, M.: A comprehensive survey on fitness landscape analysis. Recent advances in intelligent engineering systems pp. 161–191 (2012). https://doi.org/10.1007/978-3-642-23229-9_8
22. Pitzer, E., Affenzeller, M., Beham, A.: A closer look down the basins of attraction. In: 2010 UK workshop on computational intelligence (UKCI). pp. 1–6. IEEE (2010). https://doi.org/10.1109/UKCI.2010.5625595
23. Rosé, H., Ebeling, W., Asselmeyer, T.: The density of states—a measure of the difficulty of optimisation problems. In: Parallel Problem Solving from Nature—PPSN IV: International Conference on Evolutionary Computation—The 4th International Conference on Parallel Problem Solving from Nature Berlin, Germany, September 22–26, 1996 Proceedings 4. pp. 208–217. Springer (1996). https://doi.org/10.1007/3-540-61723-X_985
24. Sheikhtaheri, A., Sadoughi, F., Hashemi Dehaghi, Z.: Developing and using expert systems and neural networks in medicine: a review on benefits and challenges. Journal of medical systems **38**, 1–6 (2014). https://doi.org/10.1007/s10916-014-0110-5
25. Siems, J., Zimmer, L., Zela, A., Lukasik, J., Keuper, M., Hutter, F.: Nas-bench-301 and the case for surrogate benchmarks for neural architecture search. arXiv preprint arXiv:2008.09777 **4**, 14 (2020)
26. Stanley, K.O., Clune, J., Lehman, J., Miikkulainen, R.: Designing neural networks through neuroevolution. Nature Machine Intelligence **1**(1), 24–35 (2019). https://doi.org/10.1038/s42256-018-0006-z
27. Szegedy, C., Toshev, A., Erhan, D.: Deep neural networks for object detection. Advances in neural information processing systems **26** (2013)

28. Tavares, J., Pereira, F.B., Costa, E.: Multidimensional knapsack problem: A fitness landscape analysis. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **38**(3), 604–616 (2008). https://doi.org/10.1109/TSMCB.2008.915539
29. Thomson, S.L., Ochoa, G., Veerapen, N., Michalak, K.: Channel configuration for neural architecture: Insights from the search space. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 1267–1275 (2023). https://doi.org/10.1145/3583131.3590386
30. Traoré, K.R., Camero, A., Zhu, X.X.: Fitness landscape footprint: A framework to compare neural architecture search problems. arXiv preprint arXiv:2111.01584 (2021). https://doi.org/10.48550/arXiv.2111.01584
31. Wagner, A.: Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society B: Biological Sciences **275**(1630), 91–100 (2008). https://doi.org/10.1098/rspb.2007.1137
32. Weinberger, E.: Correlated and uncorrelated fitness landscapes and how to tell the difference. Biological cybernetics **63**(5), 325–336 (1990). https://doi.org/10.1007/BF00202749
33. White, C., Safari, M., Sukthanker, R., Ru, B., Elsken, T., Zela, A., Dey, D., Hutter, F.: Neural architecture search: Insights from 1000 papers. arXiv preprint arXiv:2301.08727 (2023). https://doi.org/10.48550/arXiv.2301.08727
34. Wright, S., et al.: The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the Sixth International Congress of Genetics (1932)
35. Ying, C., Klein, A., Christiansen, E., Real, E., Murphy, K., Hutter, F.: Nas-bench-101: Towards reproducible neural architecture search. In: International conference on machine learning. pp. 7105–7114. PMLR (2019)
36. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578 (2016). https://doi.org/10.48550/arXiv.1611.01578