ZiCo-BC: A Bias Corrected Zero-Shot NAS for Vision Tasks

Kartikeya Bhardwaj*, Hsin-Pai Cheng*, Sweta Priyadarshi*, Zhuojin Li
Qualcomm AI Research†
San Diego, CA, USA 92121
{kbhardwa, hsinpaic, swetpriy, zhuoli}@qti.qualcomm.com

Abstract

Zero-Shot Neural Architecture Search (NAS) approaches propose novel training-free metrics called zero-shot proxies to substantially reduce the search time compared to the traditional training-based NAS. Despite the success on image classification, the effectiveness of zero-shot proxies is rarely evaluated on complex vision tasks such as semantic segmentation and object detection. Moreover, existing zero-shot proxies are shown to be biased towards certain model characteristics which restricts their broad applicability. In this paper, we empirically study the bias of state-of-the-art (SOTA) zero-shot proxy ZiCo across multiple vision tasks and observe that ZiCo is biased towards thinner and deeper networks, leading to sub-optimal architectures. To solve the problem, we propose a novel bias correction on ZiCo, called ZiCo-BC. Our extensive experiments across various vision tasks (image classification, object detection and semantic segmentation) show that our approach can successfully search for architectures with higher accuracy and significantly lower latency on Samsung Galaxy S10 devices.

1. Introduction

Neural Architecture Search (NAS) algorithms have been widely used to automatically design highly accurate and efficient model architectures within a given search space. However, such techniques can be very computationally expensive as they require a lot of training resources. To address this limitation, Zero-Shot (Training-Free) NAS [7, 8, 13, 1, 16] has emerged recently, which relies on certain properties of neural network architectures to rank various models during the search without any actual training. As a result, these methods significantly accelerate the model searching process, enabling the identification of high-performing models more efficiently [7, 8, 2, 1, 16].

In this paper, we investigate two significant aspects of zero-shot NAS research. Firstly, despite the abundance of results in tasks such as image classification and various NAS-Benches [21, 11, 5, 15], several existing training-

---

*Equal Contribution. †Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.
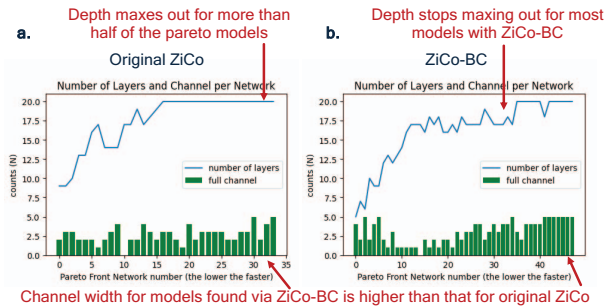

Figure 1: Overview: Zero-Shot NAS on ImageNet for EfficientNet type networks. (a) ZiCo found architectures saturate with depth and have lower channel widths, thus showing that ZiCo is biased towards thinner and deeper networks. (b) Bias-Corrected ZiCo-BC metric significantly reduces the depth-width bias and produces better models.

free metrics lack adequate validation on complex vision tasks, including semantic segmentation or object detection [7, 8, 14, 20]. Secondly, training-free metrics can be biased towards specific characteristics in neural architectures [6]. For instance, existing zero-shot NAS proxies can exhibit bias towards various factors, such as cell sizes, skip connections, convolutions, number of parameters, etc. [6].

To address the limitations mentioned above, we explore the following **key questions** centered around a recently introduced training-free metric known as ZiCo (Zero-Shot metric based on Inverse Coefficient of Variation on gradients) [7] which has demonstrated state-of-the-art performance across various NAS-Benches and ImageNet task:

1. Can ZiCo effectively perform *direct search* in complex vision tasks, such as semantic segmentation or object detection without relying on initial ImageNet search?

2. Are there any biases present in ZiCo? If yes, how can we correct these biases?

Our study demonstrates that ZiCo yields exceptional results when applied to challenging semantic segmentation and object detection tasks, especially for *macro-*

architecture search. However, when conducting *micro-architecture search* with a fixed backbone [12]), we observe a bias towards thinner (i.e., lower channel width) and deeper networks in ZiCo. Fig. 1 demonstrates zero-shot NAS results for ImageNet on a broader search space than that considered in [7]. As evident, the original ZiCo score tends to favor architectures with maximum depth and lower widths, leading to a bias towards thinner and deeper networks. This bias can hinder the effectiveness of zero-shot NAS methods across various applications, as it may lead to sub-optimal neural networks with lower accuracy. Therfore, there is a need for bias correction methods that can significantly improve the performance of zero-shot NAS.

In summary, we make the following **key contributions**: (1) We demonstrate that gradient-based zero-shot proxies like ZiCo are capable of performing *direct* macro-architecture searches on complex vision tasks such as semantic segmentation and object detection. (2) We propose a new bias correction method for ZiCo, called ZiCo-BC, that significantly enhances the metric's performance and identifies effective models in micro-architecture searches. (3) Finally, we also provide general guidelines on how to scale up this bias correction for ZiCo prior to training individual models, along with an assessment of its current limitations.

2. Zero-Shot NAS for Complex Vision Tasks

Preliminaries. ZiCo [7] is a zero-shot NAS metric that leverages the inverse coefficient of variation on gradients. This score is used to rank neural network models based on their convergence rate and generalization capacity. Specifically, ZiCo is computed as follows [7]:

$$\text{ZiCo} = \sum_{l=1}^{D} \log \left(\sum_{\theta_l} \frac{\mathbb{E}[\boldsymbol{\nabla}_{\theta_l}]}{\sqrt{\text{Var}(\boldsymbol{\nabla}_{\theta_l})}} \right), \qquad (1)$$

where, D is the total number of layers in the network, θ_l represents each parameter in layer $l \in \{1, 2, 3, \ldots, D\}$, and $\boldsymbol{\nabla}_{\theta_l}$ is the gradient of the loss w.r.t. each parameter θ_l. The expected value and standard deviation is computed across multiple batches of input data at initialization. That is, no parameters are updated across batches, only forward and backward passes are used to compute gradient statistics. It was theoretically shown in [7] that these gradient statistics are linked to training convergence and generalization.

In this paper, we will discuss two kinds of zero-shot NAS paradigms using ZiCo: **(1) Macro-Architecture Search** where we search over multiple types of backbones and heads; we call this macro-architecture search since it significantly impacts the topology of neural architectures, and **(2) Micro-Architecture Search** where the backbone type is fixed and a same type of block repeats throughout the network; here, we search over channel counts, number of block repeats, kernel sizes, type of convolution (regular, depthwise, group), expansion ratios, etc. [12].

Table 1: Direct **Macro-Architecture Search** via ZiCo on Cityscapes Semantic Segmentation

Model	#Params	#MACs	Latency (ms)	mIoU
HRNet [19]	**3.94M**	**77.89G**	28.80 (1×)	77.0%
FFNet [10]	27.49M	96.37G	**8.35** (3.4×↓)	79.7% (+2.7%)
ZiCo model	31.92M	96.14G	**8.48** (3.4×↓)	**80.7%** (+3.7%)

Does ZiCo work on complex vision tasks like semantic segmentation? A Direct Macro-Architecture Search. Despite extensive theoretical contributions and empirical validation across several NAS-Benches and ImageNet, the effectiveness of ZiCo was not evaluated for *direct search* over downstream computer vision tasks, i.e., without any prior ImageNet search. Hence, in this section, we exploit ZiCo to directly search for hardware-efficient Semantic Segmentation networks in a wide search space containing multiple types of backbones and segmentation heads.

We construct a complex search space using backbone and head from HRNet [19] architecture as well as using backbones and heads from a recent *manually-designed* hardware-efficient semantic segmentation network called FFNet [10]. HRNet [19] and FFNet [10] are highly different architectures. We also searched for HRNet-Head [19] or the Up-B-Head from FFNet [10] for head search. Finally, we introduced individual options for each backbone (e.g., depth, width, etc.), leading to a large search space.

Our objective is to exploit ZiCo to automatically design a significantly better network than the manual FFNet which was designed for mobile-scale AI accelerators. To this end, we consider the Cityscapes segmentation task and conduct NSGA-2 evolutionary search [4] over the above search space with hardware latency in the loop on the Samsung Galaxy S10 mobile platform. For ZiCo computation, we used the same loss as the one used to train FFNet [10].

Table 1 demonstrates the search results. As evident, even though the HRNet architecture has the least number of parameters and MACs, the HRNet backbone is not friendly to constrained mobile-scale hardware and shows about 3.4× higher latency compared to the manual FFNet and our automatically found ZiCo-based model, both achieving much higher accuracy. Clearly, the ZiCo-based model significantly outperforms the manual FFNet by 1% higher mIoU on Cityscapes segmentation with a similar latency.

3. Proposed Bias Correction

As mentioned earlier, we observed a bias in ZiCo towards thinner and deeper networks for micro-architecture search, i.e., when similar blocks repeat themselves throughout the fixed backbone. More precisely, in equation (1), the metric sums over the number of layers in the network. Consequently, the score grows linearly in number of layers, whereas the gradient statistics grow logarithmically. For networks with repeating blocks, this can lead to deeper

models achieving higher ZiCo scores even if they have significantly lower width. However, thinner and deeper networks may not always achieve optimal accuracy. As shown in Bhardwaj et al. [2], width plays a fundamental role in model's expressive power, and due to bias towards thinner and deeper networks, zero-shot metrics can become less effective at identifying optimal architectures. Therefore, due to the bias, ZiCo can favor deeper and thinner models over potentially more optimal ones during the evolutionary search. In the rest of this paper, we will discuss how to correct this depth-width bias in ZiCo.

Bias Correction for Micro-Architecture Search. To rectify the bias in ZiCo or other training-free NAS metrics that may exhibit a preference for thinner and deeper networks, we introduce a bias correction term. This term can be applied to modify the original metric definition. The proposed bias correction equation takes into account the *feature map resolution* and *channel width* of the network at different layers. For ZiCo, the equation is as follows:

$$
\begin{aligned}
\text{ZiCo-BC} &= \sum_{l=1}^{D} \log \left(\left[\frac{H_l W_l}{\sqrt{C_l}} \right]^{-\beta} \sum_{\theta_l} \frac{\mathbb{E}[\boldsymbol{\nabla}_{\theta_l}]}{\sqrt{\text{Var}(\boldsymbol{\nabla}_{\theta_l})}} \right) \\
&= \sum_{l=1}^{D} \log \left(\sum_{\theta_l} \frac{\mathbb{E}[\boldsymbol{\nabla}_{\theta_l}]}{\sqrt{\text{Var}(\boldsymbol{\nabla}_{\theta_l})}} \right) - \beta \sum_{l=1}^{D} \log \left(\frac{H_l W_l}{\sqrt{C_l}} \right) \\
&= \text{ZiCo} - \beta \sum_{l=1}^{D} \log \left(\frac{H_l W_l}{\sqrt{C_l}} \right)
\end{aligned}
\tag{2}
$$

Here, H_l, W_l, C_l are height, width of the feature map, and number of channels in layer l, respectively. Hyperparameter β controls the amount of depth-width penalty applied to the score. Setting $\beta = 0$ automatically yields the original ZiCo score. Clearly, if the model becomes deeper or if it has fewer channels, the penalty increases, thus, discouraging thinner and deeper models during the evolutionary search. Of note, other bias correction methods may be possible. We comment on this briefly in Section 5.

4. Experiments

We first conduct a NATS-Bench-SSS study on CIFAR-10, CIFAR-100, and ImageNet-16-120 datasets [3] to evaluate the correlations of the proposed ZiCo-BC score with accuracy and compare them to the original ZiCo score. We then evaluate the proposed bias correction for three computer vision applications: (1) ImageNet Image Classification, (2) MS COCO Object Detection, and (3) Cityscapes Semantic Segmentation. We use ResNet-based search space for semantic segmentation, and EfficientNet-based search space for ImageNet image classification as well as object detection. Next, we present more details on the micro-architecture search space and evolutionary search settings as well as performance of ZiCo-BC for each task.

Table 2: Correlation Coefficients on NATS-Bench-SSS

Dataset	Cifar-10		Cifar-100		Img16-120	
Corr. Proxy	KT	SPR	KT	SPR	KT	SPR
ZiCo	0.72	0.91	0.56	0.76	0.73	0.90
ZiCo-BC	**0.78**	**0.94**	**0.60**	**0.79**	**0.79**	**0.94**

4.1. NAS-Bench Correlations

Firstly, we evaluate the proposed bias correction on NAS benchmark NATS-Bench [11]. Specifically, we focus on the 32768 neural architectures with varying channel sizes from "size search space" (NATS-Bench-SSS), which resembles our micro-architecture search setting. Following the experimental setup in ZiCo, we compute the correlation coefficients (i.e., Kendall's τ and Spearman's ρ) between the zero-shot proxy and the test accuracy. As evident from Table 2, the bias correction improves the correlation score of ZiCo across all three datasets, indicating that the ZiCo-BC score can be a more representative proxy of test accuracy for ranking candidates during a micro-architecture search.

4.2. Classification and Object Detection

We conduct classification and object detection tasks on EfficientNet and EfficientDet, respectively [17, 18]. As these two networks share very similar backbones, we build the search space defined in previous studies [12, 17]. Specifically, the search space includes: (1) kernel size (3×3 or 5×5), (2) channel size, (3) the number of operation repeats per block, and (4) regular convolution or group convolution (with group size of 32). It is worth noting that one significant difference in our search space, as compared to existing works, is the omission of the squeeze-and-excite operation as it is not very hardware-friendly. When incorporating ZiCo-BC into the search process, we utilize the widely employed cross-entropy loss [17] for classification and the focal loss [18] for object detection, respectively.

Classification. EfficientNet [17] has proven to be a powerful architecture, achieving remarkable results in various computer vision tasks. To showcase the effectiveness of our proposed method, we employ ZiCo-BC to conduct a search for EfficientNet style models on the challenging ImageNet-1k dataset. By applying ZiCo-BC to this architecture, we aim to further enhance its performance and explore architectures that strike a balance between model depth and width. The model discovered by ZiCo-BC achieves an impressive 11% reduction in latency without sacrificing accuracy, as demonstrated in Table 3. In contrast, the original ZiCo score loses about 0.9% accuracy for similar latency.

Object Detection. EfficientDet [18] is a family of architectures renowned for high accuracy and efficiency in object detection, accommodating various resource constraints. The EfficientDet-D0 architecture comprises three

Table 3: Direct **Micro-Architecture Search** via ZiCo and ZiCo-BC on EfficientNet/Det search space

Model	Approach	Latency (ms)	Accuracy/mAP
EfficientNet	Scaling	0.90	77.7%
	ZiCo	0.82(-8%)	76.8% (-0.9%)
	ZiCo-BC	**0.80** (-11%)	**77.7%** (**0**%)
EfficientDet	Scaling	2.792	33.6
	ZiCo-BC	**1.974** (-29%)	33.8 (**+0.2%**)

Table 4: Direct **Micro-Architecture Search** via ZiCo and ZiCo-BC on Cityscapes Semantic Segmentation

Model	#Params	#MACs	Latency (ms)	mIoU
FFNet	27.49M	96.37G	8.35	79.70%
ZiCo	**21.80M**	**75.89G**	**7.02** (-16%)	78.62% (-1.08%)
ZiCo-BC	23.28M	79.85G	7.44 (-11%)	**79.71%** (**+0.01%**)

components: (1) an EfficientNet backbone network, (2) a weighted bi-directional feature pyramid network (BiFPN), and (3) a class and box network for predicting object class and bounding box information. Notably, the backbone of EfficientDet-D0 contributes 78% of the FLOPs and 92% of the parameters in the entire architecture. Hence, our primary focus lies in searching for a backbone network that enhances the latency of the architecture without sacrificing accuracy. Table 3 displays the results on MS COCO 2017 [9] after training for 300 epochs. Our searched architecture achieves a remarkable 29% latency reduction while maintaining even better accuracy compared to EfficientDet-D0.

4.3. Semantic Segmentation

In this section, we evaluate the bias correction ability of the proposed ZiCo-BC score in the context of micro-architecture search on Cityscapes dataset. Unlike Section 2, where we conducted a macro-architecture search across HRNet [19] and FFNet [10], here we specifically test ZiCo-BC on the FFNet backbone in conjunction with the FFNet-Head. The FFNet backbone is based on the ResNet architecture and consists of four stages. Our micro-architecture search space consists of (1) number of residual blocks in each stage, (2) number of output channels for each stage; each residual block in the stage has the same number of channels, and (3) type of convolution, i.e., Group Convolution with a group size of 32/64/128 channels whichever is larger, or a Regular Convolution. All kernel sizes are fixed to 3×3. To search over a large space, we significantly vary the width and depth of the candidate networks around the baseline FFNet [10] configuration. Overall, this search space consists of more than 44M unique architectures.

Table 4 shows that ZiCo-BC finds a model with similar mIoU as FFNet [10] but achieves 11% lower latency on the mobile platform. In contrast, networks found via the original ZiCo metric lose nearly 1% mIoU with about 16% lower latency. The ZiCo-BC model has 74 residual blocks with higher channel widths, thus, correcting the bias towards deeper and thinner networks. Improving the latency of FFNet [10] by 11% (with similar mIoU) is highly nontrivial as it is already designed for mobile devices.

5. General Guidelines and Limitations

General Guidelines. One crucial aspect of the proposed bias correction is determining the value of the main hy-

perparameter β, which influences the penalty on depth and width. An appropriate β value can be obtained by analyzing the architecture of Pareto models found during evolutionary search. For instance, in Fig. 1(a), most models exhibit maximum depth with low width, indicating the presence of bias. To address this, we gradually increase β to encourage more diverse architectures with intermediate depth. For classification and object detection tasks, we used $\beta = 1$. On the other hand, we used $\beta = 2$ for semantic segmentation zero-shot micro-architecture search.

Limitations. Two limitations of our current bias correction are identified. Firstly, the bias correction applies solely to micro-architecture search with repeated blocks. For macro-architecture search, gradient statistics for different backbones (topologies) result in a similar score even if they have highly different number of layers[1]. Therefore, a common penalty to backbones with different depths would treat the shallower backbone unfairly. Hence, further research is needed to come up with a universal bias correction (if needed) for macro-architecture search. Secondly, the current bias correction assumes a fixed input size for candidate models, disregarding the potential gain in accuracy for various vision tasks by increasing the image size. Hence, future bias correction methods that maintain the overall score with increasing input size are of interest.

6. Conclusion

In this paper, we explore the effectiveness of zero-shot NAS on complex vision tasks beyond traditional image classification. Firstly, we validate an existing proxy, called ZiCo for *macro-architecture search* in semantic segmentation. The ZiCo-based network achieves a remarkable $3.4\times$ speed up over HRNet through automatic search, with 1% higher mIoU compared to a manually designed model of similar latency. Next, we identify biases in ZiCo for *micro-architecture search* and propose ZiCo-BC, a novel bias correction method for depth-width biases in zero-shot metrics. Finally, we demonstrate that our bias correction enables ZiCo-BC to consistently achieve 11-30% lower latency and 0.2-1.1% higher accuracy compared to the models found via the original ZiCo for micro-architecture search on image classification, object detection, and segmentation.

---

[1] We observed this between FFNet and HRNet candidates: For networks with similar trainability, HRNet-based models with nearly half the layers often achieve comparable ZiCo to deeper FFNet-based models.

References

[1] Kartikeya Bhardwaj, Guihong Li, and Radu Marculescu. How does topology influence gradient propagation and model performance of deep networks with densenet-type skip connections? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13498–13507, 2021.

[2] Kartikeya Bhardwaj, James Ward, Caleb Tung, Dibakar Gope, Lingchuan Meng, Igor Fedorov, Alex Chalfin, Paul Whatmough, and Danny Loh. Restructurable activation networks. *arXiv preprint arXiv:2208.08562*, 2022.

[3] Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. A downsampled variant of imagenet as an alternative to the cifar datasets. *arXiv preprint arXiv:1707.08819*, 2017.

[4] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[5] Xuanyi Dong and Yi Yang. Nas-bench-201: Extending the scope of reproducible neural architecture search. *arXiv preprint arXiv:2001.00326*, 2020.

[6] Arjun Krishnakumar, Colin White, Arber Zela, Renbo Tu, Mahmoud Safari, and Frank Hutter. Nas-bench-suite-zero: Accelerating research on zero cost proxies. *Advances in Neural Information Processing Systems*, 35:28037–28051, 2022.

[7] Guihong Li, Yuedong Yang, Kartikeya Bhardwaj, and Radu Marculescu. ZiCo: Zero-shot NAS via inverse Coefficient of Variation on Gradients. In *The Eleventh International Conference on Learning Representations*, 2022.

[8] Ming Lin, Pichao Wang, Zhenhong Sun, Hesen Chen, Xiuyu Sun, Qi Qian, Hao Li, and Rong Jin. Zen-nas: A zero-shot nas for high-performance image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 347–356, 2021.

[9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[10] Dushyant Mehta, Andrii Skliar, Haitam Ben Yahia, Shubhankar Borse, Fatih Porikli, Amirhossein Habibian, and Tijmen Blankevoort. Simple and efficient architectures for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2628–2636, 2022.

[11] Yash Mehta, Colin White, Arber Zela, Arjun Krishnakumar, Guri Zabergja, Shakiba Moradian, Mahmoud Safari, Kaicheng Yu, and Frank Hutter. Nas-bench-suite: Nas evaluation is (now) surprisingly easy. *arXiv preprint arXiv:2201.13396*, 2022.

[12] Bert Moons, Parham Noorzad, Andrii Skliar, Giovanni Mariani, Dushyant Mehta, Chris Lott, and Tijmen Blankevoort. Distilling optimal neural networks: Rapid search in diverse spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12229–12238, 2021.

[13] Xuan Shen, Yaohua Wang, Ming Lin, Yilun Huang, Hao Tang, Xiuyu Sun, and Yanzhi Wang. Deepmad: Mathematical architecture design for deep convolutional neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6163–6173, 2023.

[14] Yao Shu, Zhongxiang Dai, Zhaoxuan Wu, and Bryan Kian Hsiang Low. Unifying and boosting gradient-based training-free neural architecture search. *Advances in Neural Information Processing Systems*, 35:33001–33015, 2022.

[15] Julien Niklas Siems, Lucas Zimmer, Arber Zela, Jovita Lukasik, Margret Keuper, and Frank Hutter. Nas-bench-301 and the case for surrogate benchmarks for neural architecture search. 2020.

[16] Zhenhong Sun, Ming Lin, Xiuyu Sun, Zhiyu Tan, Hao Li, and Rong Jin. Mae-det: Revisiting maximum entropy principle in zero-shot nas for efficient object detection. *arXiv preprint arXiv:2111.13336*, 2021.

[17] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[18] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020.

[19] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3349–3364, 2020.

[20] Lichuan Xiang, Lukasz Dudziak, Mohamed S Abdelfattah, Thomas Chau, Nicholas D Lane, and Hongkai Wen. Zero-cost operation scoring in differentiable architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10453–10463, 2023.

[21] Chris Ying, Aaron Klein, Eric Christiansen, Esteban Real, Kevin Murphy, and Frank Hutter. Nas-bench-101: Towards reproducible neural architecture search. In *International conference on machine learning*, pages 7105–7114. PMLR, 2019.