Appendix for Extensible and Efficient Proxy for Neural Architecture Search

1. Experiment Setup

Eproxy The learning rate is 1.0, and the weight decay is $1e{-5}$. Each architecture is trained for ten iterations with 16 images randomly sampled from the CIFAR-10 training set as a mini-batch (tiny dataset). The SGD optimizer is used for training.

DPS The total evolution cycle is 200. The number of architectures sampled for ranking is 20. The population size is 40. The sample size is 10. The mutation rate is 0.2.

1.1. GPU Benchmark

We benchmark the average evaluation time for architecture with Eproxy and GPU utilization on different search spaces (shown in Table 1). For DPS, it's straightforward to estimate the total time. For example, if we conduct DPS on NDS-DARTS search space with 20 architectures to get each proxy's ranking correlation and 200 total evolution cycles, the time is $\sim 20 \times 200 \times 0.72 = 2880$ seconds. All experiments are done on a single A6000 GPU.

1.2. Search Spaces

NAS-Bench-101 [14]: 423K CNN architectures are trained on CIFAR-10 dataset.

NAS-Bench-201 [3]: 15625 CNN architectures are trained on CIFAR-10/CIFAR-100/TinyImageNet.

NDS dataset [11]: **DARTS**: A DARTS [9] style search space including 5000 sampled architectures trained on CIFAR-10. **DARTS-fix_w_d**: A DARTS style search space with fixed width and depth including 5000 sampled architectures trained on CIFAR-10. **AmoebaNet**: An AmoebaNet [12] style search space including 4983 sampled architectures trained on CIFAR-10. **ENAS**: An ENAS [10] style search space including 4999 sampled architectures trained on CIFAR-10. **ENAS-fix_w_d**: An ENAS style search space with fixed width and depth including 5000 sampled architectures trained on CIFAR-10. **NASNet**: A NASNet [16] style search space including 4846 sampled architectures trained on CIFAR-10. **PNAS**: A PNAS [8] style search space including 4999 sampled architectures trained on CIFAR-10. **PNAS-fix_w_d**: A PNAS style search space with fixed width and depth including 4559 sampled architectures trained on CIFAR-10. **ResNet**: A ResNet [6] style search space including 25000 sampled architectures trained on CIFAR-10. **ResNeXt-A**: A ResNeXt [13] style search space including 24999 sampled architectures trained on CIFAR-10. **ResNeXt-B**: Another ResNeXt style search space including 25508 sampled architectures trained on CIFAR-10. **DARTS_in**: A DARTS style search space including 121 sampled architectures trained on ImageNet-1k. **DARTS-fix_w_d-in**: A DARTS style search space with fixed width and depth including 499 sampled architectures trained on ImageNet-1k. **Amoeba_in**: An AmoebaNet style search space including 124 sampled architectures trained on ImageNet-1k. **ENAS_in**: A ENAS style search space including 117 sampled architectures trained on ImageNet-1k. **NASNet_in**: A NASNet style search space including 122 sampled architectures trained on ImageNet-1k. **PNAS_in**: A PNAS style search space including 119 sampled architectures trained on ImageNet-1k. **ResNeXt-A_in**: A ResNeXt style search space including 130 sampled architectures trained on ImageNet-1k. **ResNeXt-B_in**: Another ResNeXt style search space including sampled 164 architectures trained on ImageNet-1k.

NAS-Bench-Trans-Micro [4]: A NAS-Bench-201 style search space including 4096 architectures trained on 7 different tasks on the subsets of Taskonomy dataset [15]. Tasks including: **Object Classification** for 75 classes of objects. **Scene Classification** for 47 classes of scenes. **Room Layout** for estimating and aligning a 3D bounding box by utilizing a 9-dimension vector. **Jigsaw Content Prediction** by dividing the input image into 9 patches and shuffling according to one of 1000 preset permutations. **Semantic Segmentation** for 17 semantic classes. **Autoencoding** for reconstructing the input images.

NAS-Bench-MR [2]: A complex search space for multi-resolution networks including 2507 trained architectures on 9 different tasks. Tasks including: **ImageNet-50-1000 (Cls-A)** with 50 classes and 1000 samples from each class from ImageNet-1k. **ImageNet-50-100 (Cls-B)** with 50 classes and 100 samples from each class from ImageNet-1k. **ImageNet-10-1000 (Cls-A)** with 10 classes and 1000 samples from each class from ImageNet-1k. **ImageNet-10c** same as Cls-A but architectures are trained for 10 epochs. **Seg** for Cityscapes dataset [1]. **Seg-4x** for Cityscapes dataset with 4x downsampled resolution. **3dDet** on KITTI dataset [5]. **Video** for HMDB51 dataset [7]. **Video-p** for HMDB51 but architectures are pretrained with ImageNet-

Search space	NB101	NB201	DARTS	DARTS-fix-w-d	Amoeba
Avg. Eval. Time (ms)	414.1	324.0	719.2	1198.3	1191.3
GPU Util. (MB)	4137	1603	3221	2275	3365
Search space	ENAS	ENAS-fix-w-d	NASNet	PNAS	PNAS-fix-w-d
Avg. Eval. Time (ms)	908.2	1408.2	878.7	1041.4	1824.7
GPU Util. (MB)	3245	2577	3129	3391	3447
Search space	ResNet	ResNeXt-A	ResNeXt-B	NAS-Bench-Trans-Micro	NAS-Bench-MR
Avg. Eval. Time (ms)	242.3	314.5	298.7	355.2	1011.9
GPU Util. (MB)	2765	2423	2777	2081	4229

Table 1. Average time for evaluating an architecture with Eproxy in the target search space and Maximum GPU utilization. The results suggest that Eproxy is efficient and computation-friendly.

50-1000.

1.3. Searched Architectures

The searched architectures for DARTS-ImageNet search space are shown in Fig 1.

References

[1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1

[2] Mingyu Ding, Yuqi Huo, Haoyu Lu, Linjie Yang, Zhe Wang, Zhiwu Lu, Jingdong Wang, and Ping Luo. Learning versatile neural architectures by propagating network codes. *arXiv preprint arXiv:2103.13253*, 2021. 1

[3] Xuanyi Dong and Yi Yang. Nas-bench-201: Extending the scope of reproducible neural architecture search. In *International Conference on Learning Representations (ICLR)*, 2020. 1

[4] Yawen Duan, Xin Chen, Hang Xu, Zewei Chen, Xiaodan Liang, Tong Zhang, and Zhenguo Li. Transnas-bench-101: Improving transferability and generalizability of cross-task neural architecture search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5251–5260, 2021. 1

[5] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[7] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. Hmdb: a large video database for human motion recognition. In *2011 International conference on computer vision*, pages 2556–2563. IEEE, 2011. 1

[8] Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens, Wei Hua, Li-Jia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. Progressive neural architecture search. In *Proceedings of the European conference on computer vision (ECCV)*, pages 19–34, 2018. 1

[9] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018. 1

[10] Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. Efficient neural architecture search via parameters sharing. In *International conference on machine learning*, pages 4095–4104. PMLR, 2018. 1

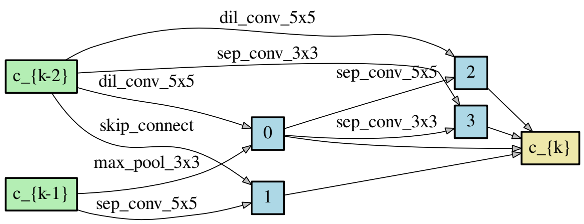
[11] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10428–10436, 2020. 1

[12] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V Le. Regularized evolution for image classifier architecture search. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, pages 4780–4789, 2019. 1

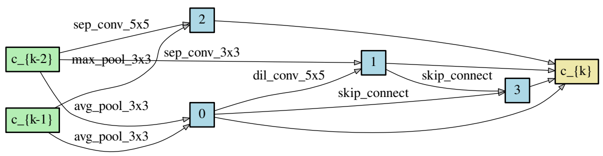
[13] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017. 1

[14] 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. 1

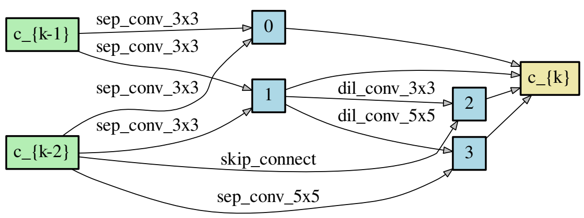
[15] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3712–3722, 2018. 1



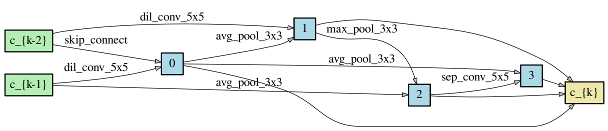
(a) Eproxy Normal Cell



(b) Eproxy Reduction Cell



(c) Eproxy+DPS Normal Cell



(d) Eproxy+DPS Reduction Cell

Figure 1. Visualize the architecture found by Eproxy and Eproxy+DPS on ImageNet-DARTS search space.

[16] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8697–8710, 2018. 1