

# Towards Improving the Consistency, Efficiency, and Flexibility of Differentiable Neural Architecture Search (Appendices)

Yibo Yang<sup>1,2</sup>, Shan You<sup>3</sup>, Hongyang Li<sup>2</sup>, Fei Wang<sup>3</sup>, Chen Qian<sup>3</sup>, Zhouchen Lin<sup>2</sup>

<sup>1</sup>Center for Data Science, Academy for Advanced Interdisciplinary Studies, Peking University

<sup>2</sup>Key Laboratory of Machine Perception (MOE), School of EECS, Peking University

<sup>3</sup>SenseTime

{ibo, lhy\_ustb, zlin}@pku.edu.cn, {youshan, wangfei, qianchen}@sensetime.com

## Appendix A: Datasets and Search Spaces

### A.1 Datasets

The CIFAR-10 dataset consists of 60,000 images of size  $32 \times 32$  in 10 classes. There are 50,000 images for training and 10,000 images for testing. In search, we use a half of the training set to optimize network weights and the other half as the validation set to optimize architecture parameters.

The ImageNet dataset contains 1.2 million training images, 50,000 validation images, and 100,000 test images. Following [3], we directly perform the search on ImageNet by randomly sampling 10% images of the training set for network weights, and another 10% for architecture parameters.

The standard data augmentation methods are used for both CIFAR-10 and ImageNet.

### A.2 Search Spaces

Following [2, 1, 3], we search on the commonly used operation space that includes  $3 \times 3$  and  $5 \times 5$  separable convolution,  $3 \times 3$  and  $5 \times 5$  dilated separable convolution,  $3 \times 3$  max pooling,  $3 \times 3$  average pooling, identity, and zero. Because our method is friendly to memory consumption and thus can be performed on a larger search space, we adopt an extended version of operation space that has 5 extra operations:  $1 \times 1$  convolution,  $3 \times 3$  convolution,  $1 \times 3$  then  $3 \times 1$  convolution,  $1 \times 5$  then  $5 \times 1$  convolution, and  $1 \times 7$  then  $7 \times 1$  convolution. The two versions of operation space are listed in Table 1. The zero operation is used to indicate the lack of connection between two nodes [2]. For EngineNAS, we keep the zero operation as convention, while we remove the zero operation for EngineNAS-DST because it inherently has the ability to learn topology in an explicit manner.

## Appendix B: Visualization of Searched Architectures

We visualize all searched architectures of our methods. Concretely, the EnTranNAS searched on the standard space

| Operation                                  | Standard Space | Extended Space |
|--|----------------|----------------|
| zero                                       | ✓              | ✓              |
| $3 \times 3$ separable convolution         | ✓              | ✓              |
| $5 \times 5$ separable convolution         | ✓              | ✓              |
| $3 \times 3$ dilated separable convolution | ✓              | ✓              |
| $5 \times 5$ dilated separable convolution | ✓              | ✓              |
| $3 \times 3$ max pooling                   | ✓              | ✓              |
| $3 \times 3$ average pooling               | ✓              | ✓              |
| identity                                   | ✓              | ✓              |
| $1 \times 1$ convolution                   | -              | ✓              |
| $3 \times 3$ convolution                   | -              | ✓              |
| $1 \times 3$ then $3 \times 1$ convolution | -              | ✓              |
| $1 \times 5$ then $5 \times 1$ convolution | -              | ✓              |
| $1 \times 7$ then $7 \times 1$ convolution | -              | ✓              |

Table 1. The standard and extended operation spaces.

of CIFAR-10 is shown in Figure 1 and 2. Its result on the extended space is shown in Figure 3 and 4. The EnTranNAS-DST searched on CIFAR-10 is shown in Figure 5 and 6. The EnTranNAS (ImageNet) directly searched on ImageNet is shown in Figure 7 and 8. The EnTranNAS-DST (ImageNet) on ImageNet is shown in Figure 9 and 10. In our ablation experiments, the results of EnTranNAS-DST with different  $\lambda$  on ImageNet are shown from Figure 11 to 16.

To better inspect the search process of EnTranNAS-DST, we record the derived architecture of each epoch in the video attached in the supplementary file. It is shown that at the beginning of search, all connections are kept active. That is because by the initialization of  $\{\alpha_k^{(i,j)}\}$ ,  $\{p_k^{(i,j)}\}$  have similar values  $\forall i < j, 1 \leq k \leq K$ , and thus  $\{\hat{p}_k^{(i,j)}\}$  are large by Eq. (7) and will not be thresholded by Eq. (8), i.e.,  $\hat{q}_k^{(i,j)} > 0, \forall i, j, k$ . As the optimization proceeds, connections begin to be cut out gradually. we finally obtain an architecture with both operation and topology learnable. The topology is not subject to any hand-crafted rule.

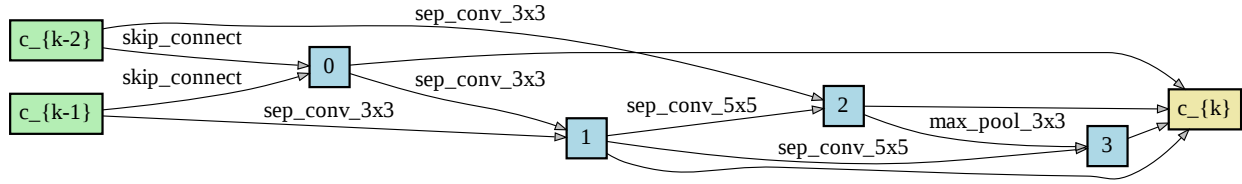


Figure 1. EnTranNAS normal cell searched on CIFAR-10 (the result in Table 6).

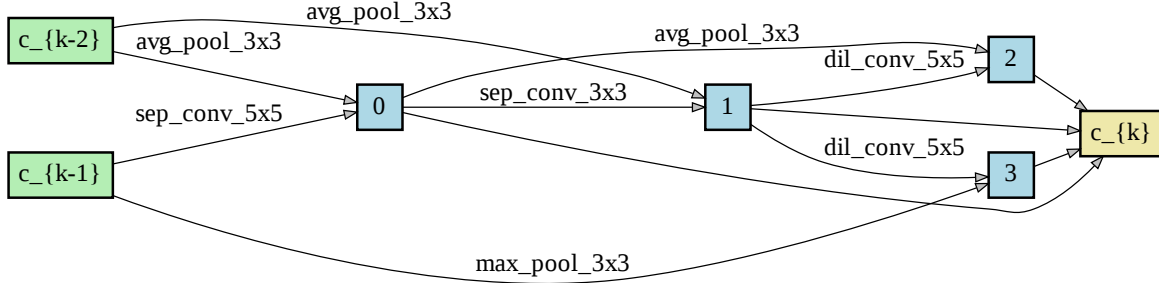


Figure 2. EnTranNAS reduction cell searched on CIFAR-10 (the result in Table 6).

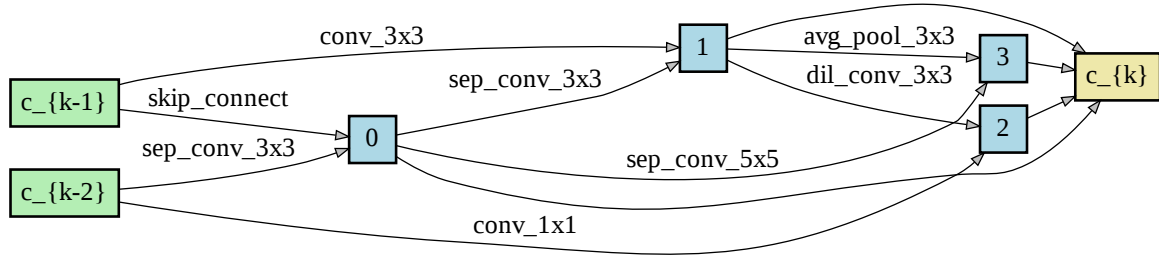


Figure 3. EnTranNAS (12 operations) normal cell searched on CIFAR-10 (the result in Table 6).

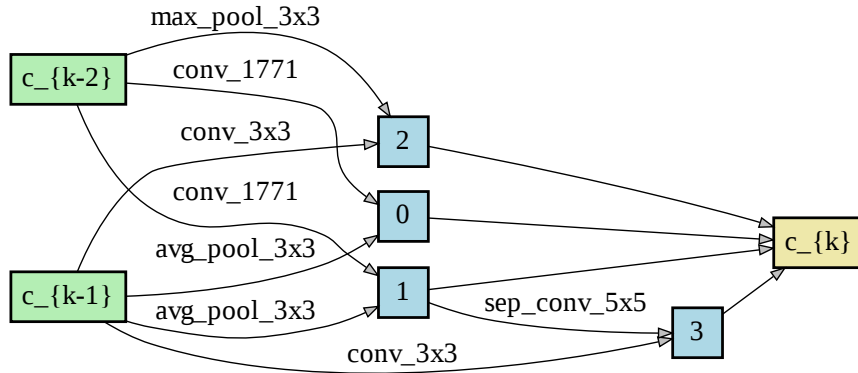


Figure 4. EnTranNAS (12 operations) reduction cell searched on CIFAR-10 (the result in Table 6).



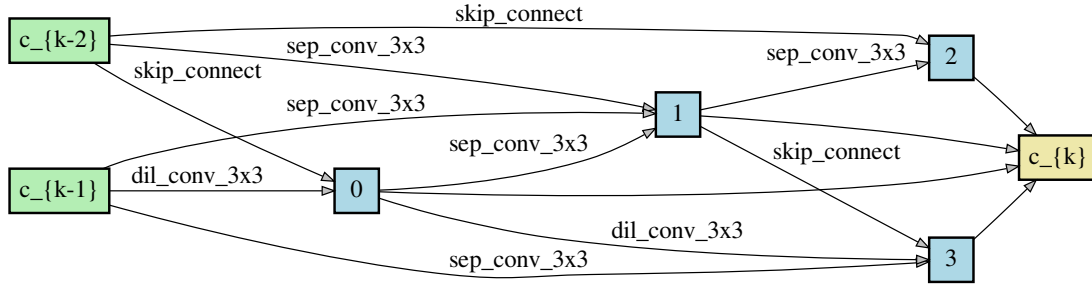


Figure 9. EnTranNAS-DST (ImageNet) normal cell searched on CIFAR-10 (the result in Table 7).

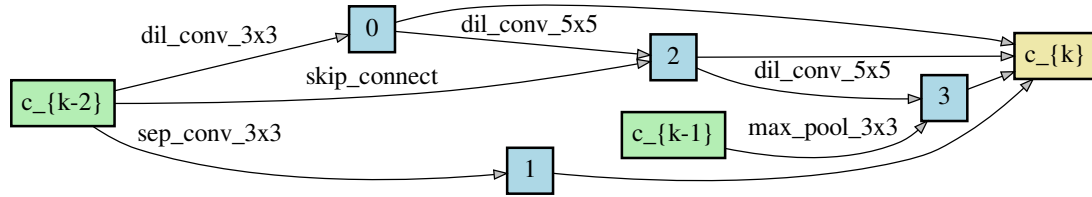


Figure 10. EnTranNAS-DST (ImageNet) reduction cell searched on CIFAR-10 (the result in Table 7).

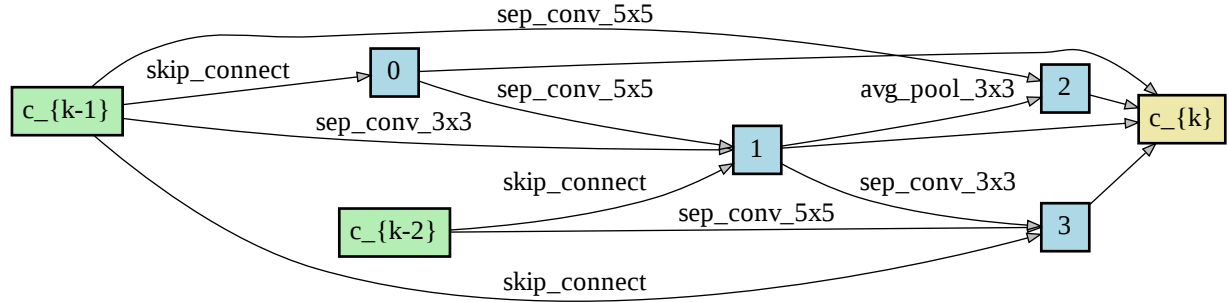


Figure 11. EnTranNAS-DST ( $\lambda = 0.2$ ) normal cell searched on ImageNet (the result in Table 4).

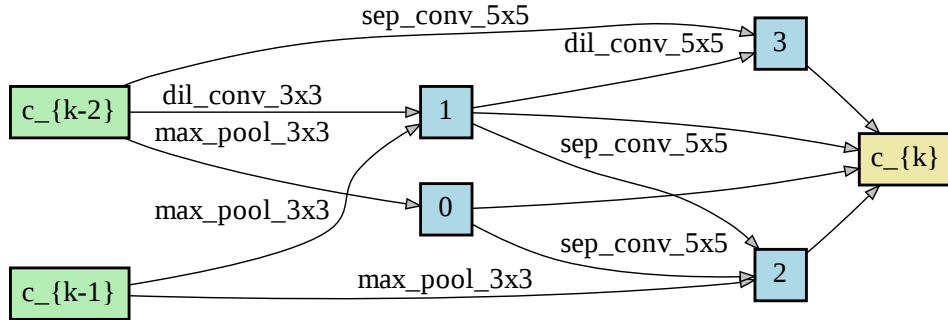


Figure 12. EnTranNAS-DST ( $\lambda = 0.2$ ) reduction cell searched on ImageNet (the result in Table 4).

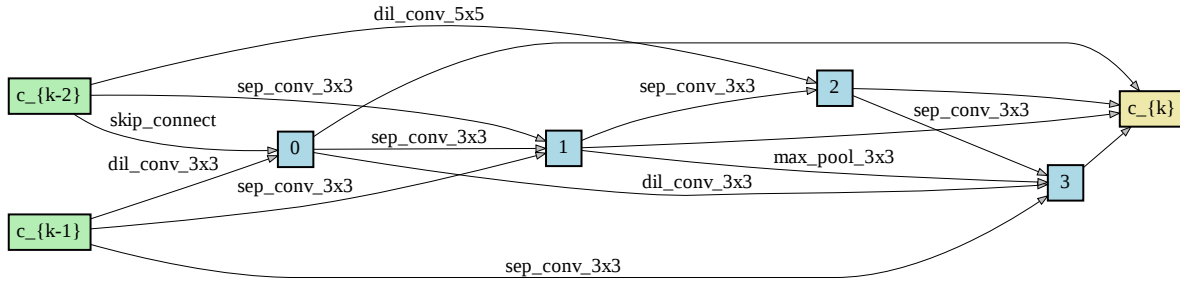


Figure 13. EnTranNAS-DST ( $\lambda = 0.1$ ) normal cell searched on ImageNet (the result in Table 4).

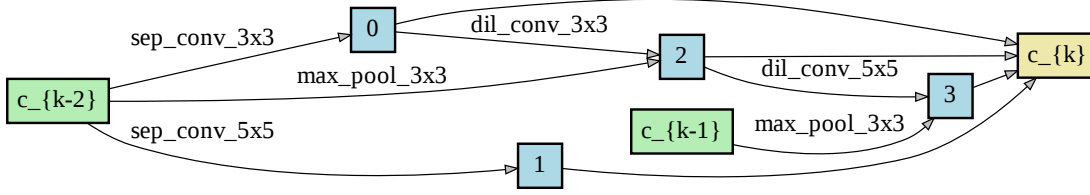


Figure 14. EnTranNAS-DST ( $\lambda = 0.1$ ) reduction cell searched on ImageNet (the result in Table 4).

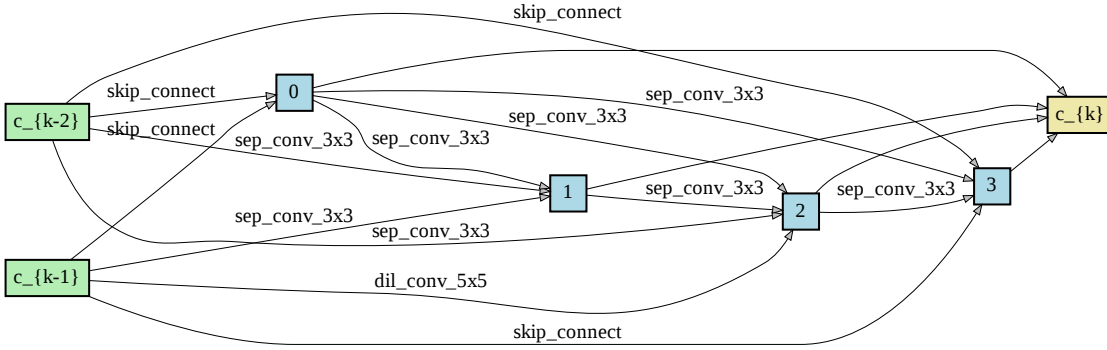


Figure 15. EnTranNAS-DST ( $\lambda = 0.05$ ) normal cell searched on ImageNet (the result in Table 4).

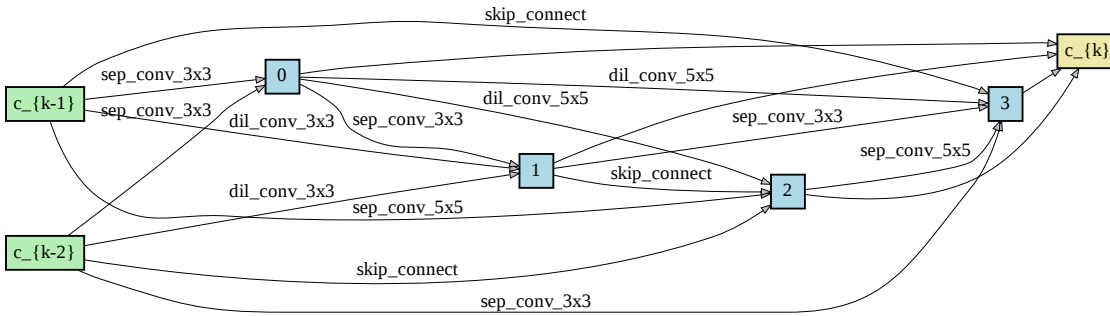


Figure 16. EnTranNAS-DST ( $\lambda = 0.05$ ) reduction cell searched on ImageNet (the result in Table 4).

## References

- [1] Xin Chen, Lingxi Xie, Jun Wu, and Qi Tian. Progressive differentiable architecture search: Bridging the depth gap between search and evaluation. In *ICCV*, pages 1294–1303, 2019. [1](#)
- [2] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. In *ICLR*, 2019. [1](#)
- [3] Yuhui Xu, Lingxi Xie, Xiaopeng Zhang, Xin Chen, Guo-Jun Qi, Qi Tian, and Hongkai Xiong. Pc-darts: Partial channel connections for memory-efficient differentiable architecture search. In *ICLR*, 2020. [1](#)