Revisiting Training-free NAS Metrics: An Efficient Training-based Method - Supplementary

Taojiannan Yang[1]*, Linjie Yang[2], Xiaojie Jin[2], Chen Chen[1]
[1]Center for Research in Computer Vision, University of Central Florida [2]Bytedance Inc.
taoyang1122@knights.ucf.edu {linjie.yang, jinxiaojie}@bytedance.com
chen.chen@crcv.ucf.edu

1. Implementation Details

As explained in Section 4 in the paper, we randomly sample 10 classes and 10 images from each class in each run. The network is trained for 50 iterations with a fixed learning rate of 0.2. We use the same sampling strategy on all datasets (CIFAR-10/100, ImageNet16-120 and ImageNet-1K). On NAS-Bench-201 pruning-based method, we increase the training iterations to 100 because it takes longer for the supernet to converge. On DARTS search space, we increase the learning rate to 0.5 for better convergence. For evaluation, we strictly follow the settings in TE-NAS.

2. Searched Architectures

We visualize the searched architectures on DARTS search space using our metrics (*AngleLoss* and *AngleLoss+#Param*) in Fig. 3 to Fig. 6. We show the architectures searched directly on CIFAR-10 and ImagaNet, respectively.

3. More explanations on the trivial structures

As shown in Section 5 of the paper, our metric (AngleLoss) does not perform well on the overall search space. The reason is that the *Angle* metric may give a very high score to trivial structures where most of the connections are *zeroize, skip-connect* or *avg_pool*. Our conjecture is that in these trivial structures, the feature extraction layers do not learn anything meaningful, and the prediction layer is optimized towards random directions in each training iteration. So the weight vector of the prediction layer almost does not change after training, which means $Angle$ metric will give a high score to these structures. Here we show an example of the trivial structure and how it is resolved when *Angle* metric is combined with #Param. Fig. 1 shows the rank index and accuracy of different structures on NAS-Bench-201 CIFAR-100. We randomly select 100 networks and rank them using *Angle* metric. The orange dot in Fig. 1 has the highest score. However, its accuracy is very low. We visualize its cell structure on the right. We find that the orange dot is a trivial structure which is only composed of *zeroize, skip-connect* and *avg_pool*. As explained above, *Angle* metric may give a high score to such structures, therefore the performance on the overall search space is not good and has a very large variance. However, such trivial structures can be easily removed when *Angle* metric is combined with #Param. In Fig. 2, we show the rank index of the combined metric on the same set of structures. The trivial structure (orange dot) is now ranked around the middle of all networks because it has a very low rank index in terms of #Param. As a result, #Param could help remove the trivial solution improperly discovered by *Angle* metric.

4. Correlation of different metrics with test accuracy

We show the Kendall's Tau of different metrics with test accuracy on NAS-Bench-201 in Tab. 1. We evaluate Kendall's Tau over three groups of networks, all networks (15625 networks) in NAS-Bench-201, networks whose test accuracy is within top-1000 and networks whose test accuracy is within top-500. As shown in Tab .1, #Param has a good correlation with test accuracy. LR2 has the highest correlation on the overall search space. However, as discussed in the paper, LR2 has a very high correlation with #Param so it is not surprising that it is good in this case. AngleLoss does not have a very high correlation, but it may collapse to some trivial networks as explained in Sec. 3. When combined with the #Param, the correlation score of AngleLoss is significantly boosted and is comparable with other metrics. This again validates that our metric provides additional information in ranking networks which can be effectively combined with the #Param.

We also show the correlation on the Top-1000 and Top-500 networks, since in practice we may be more interested in finding the best networks from a group of comparable

*Work done during an internship at Bytedance Inc.

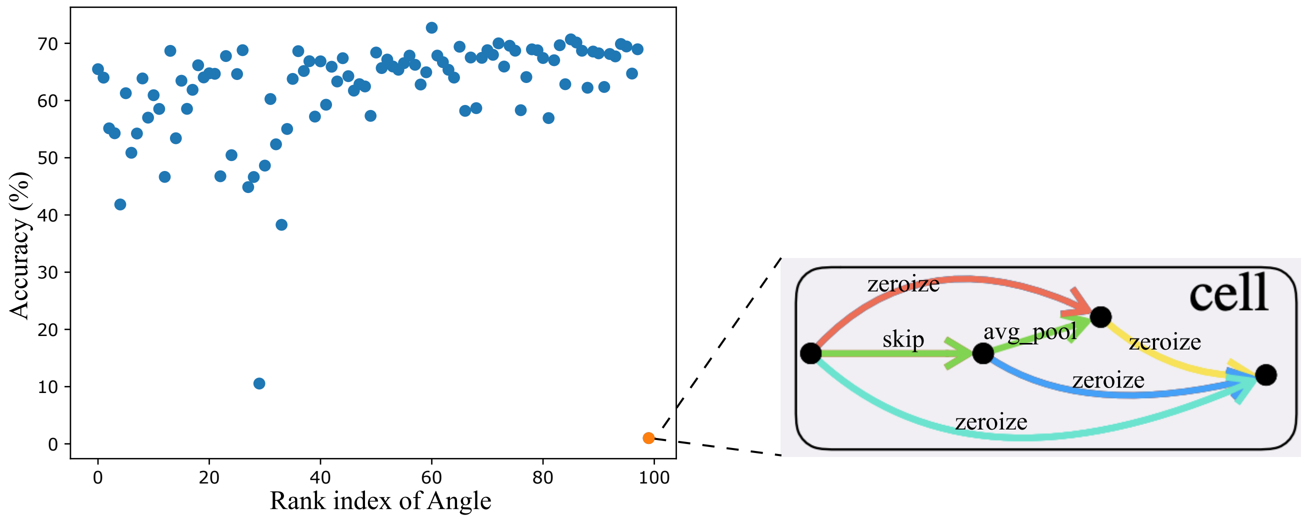


Figure 1: Rank index of different structures using *Angle* metric. The orange dot has the highest score. Its structure is visualized on the right. This is based on NAS-Bench-201 CIFAR-100 dataset and random search.

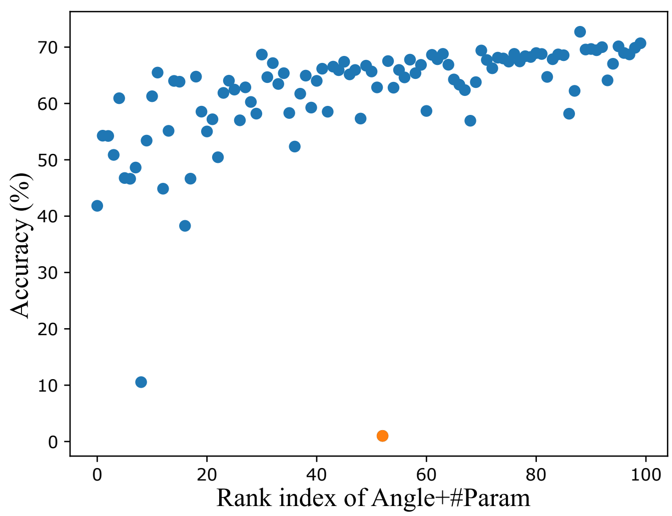
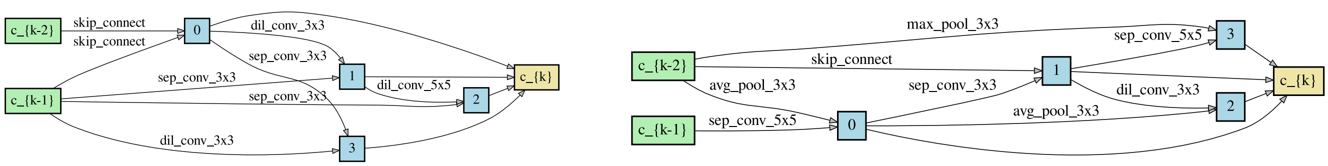


Figure 2: Rank index of different structures using *Angle+#Param*. The orange dot is the one that has the highest score in Fig. 1.

networks. A metric could achieve a good ranking score by doing well in low-performance networks but not very well in high-performance networks. In Tab. 1, we can see that the #Param is not good in high-performance networks. Similarly, the performance of LR2 drops dramatically to around 0, which means LR2 can not distinguish the high-performance networks. The performance of other training-free metrics also drops a lot. However, in this scenario, our metric is significantly better than training-free metrics. This shows that our metric is better in ranking high-performance networks.

Table 1: kendall's Tau of different metrics with test accuracy on NAS-Bench-201. Top-1000 means the Kendall's Tau on networks whose test accuracy is in top-1000 in all 15625 networks.

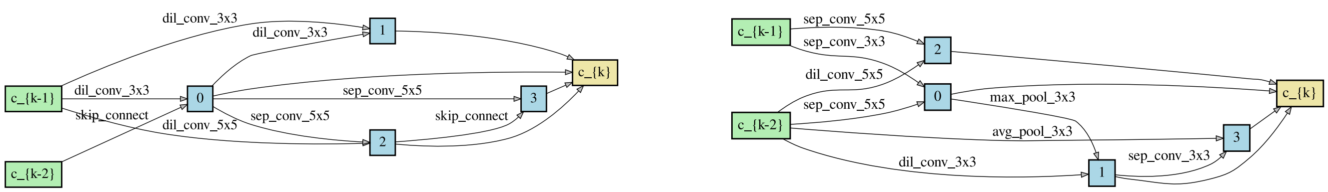
Metrics	All 15625 networks	Top-1000	Top-500
#Param	0.55	0.04	0.03
LR1	0.50	0.15	0.08
NTK	0.42	0.06	0.14
LR2	**0.61**	-0.04	-0.03
AngleLoss	0.46	**0.15**	**0.18**
LR1+#Param	0.62	0.15	0.10
NTK+#Param	0.56	0.07	0.14
LR2+#Param	**0.64**	0.00	0.02
AngleLoss+#Param	0.60	**0.17**	**0.21**



Normal Cell

Reduction Cell

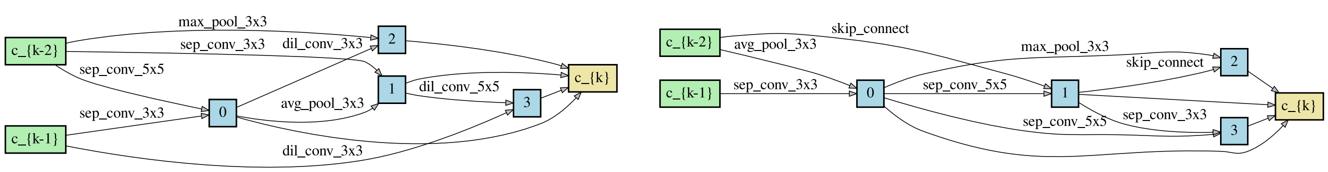
Figure 3: Normal and Reduction cells discovered by *AngleLoss* metric on DARTS CIFAR-10.



Normal Cell

Reduction Cell

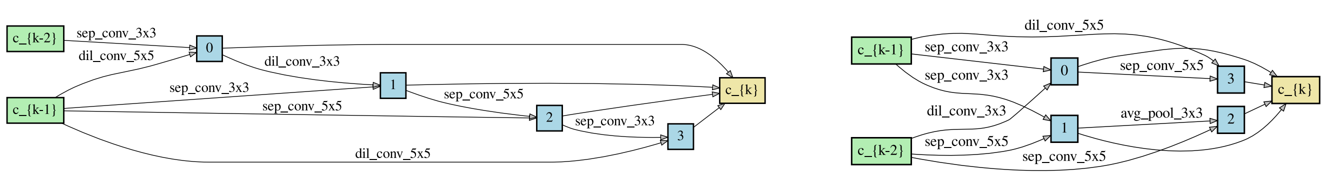
Figure 4: Normal and Reduction cells discovered by *AngleLoss+#Param* metric on DARTS CIFAR-10.



Normal Cell

Reduction Cell

Figure 5: Normal and Reduction cells discovered by *AngleLoss* metric on DARTS ImageNet.



Normal Cell

Reduction Cell

Figure 6: Normal and Reduction cells discovered by *AngleLoss+#Param* metric on DARTS ImageNet.