Supplementary Material for Curriculum-NAS: Curriculum Weight-Sharing Neural Architecture Search

ABSTRACT

In this supplementary material, we provide detailed information of the pre-experiment on the influence of data on architectures in Section 1 of the main paper. Besides, we present the changing of data weights in the searching procedure of Curriculum-NAS in Section 2.

1 INFLUENCE OF DATA ON ARCHITECTURES

In Section 1 of the main paper, we illustrate the misclassification similarity of 10 networks on the training and test set of CIFAR-10. We now present the detailed information of this pre-experiment.

The former 5 common networks are MobileNet[4], VGG[5], ResNet[1], DenseNet[2] and DLA[6], all of which are designed for image classification. Their architectures are described in detail in their respective papers.

The later 5 networks are sampled randomly from DARTS search space[3] with the fixed seed 666. Their architectures are as follows:

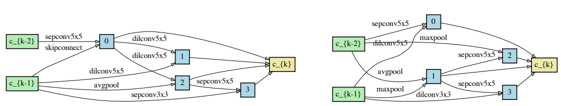


Figure 1: Normal (left) and reduction (right) cell of subnet1

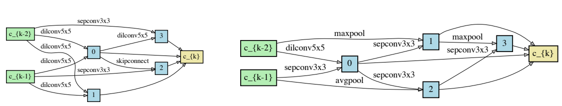


Figure 2: Normal (left) and reduction (right) cell of subnet2

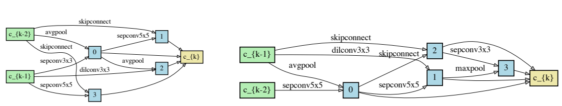


Figure 3: Normal (left) and reduction (right) cell of subnet3

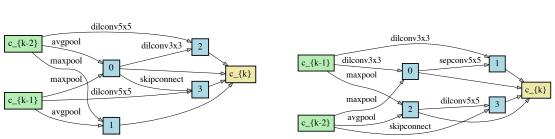


Figure 4: Normal (left) and reduction (right) cell of subnet4

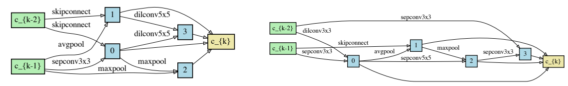


Figure 5: Normal (left) and reduction (right) cell of subnet5

We follow the training framework of Pytorch-Cifar[1] to train and evaluate their performances. Table 1 shows the training and test accuracies of the 10 networks and verifies that all of them are trained to convergence.

Table 1: Image classification Accuracies(%)

	Train Acc(%)	Test Acc(%)
MobileNet	98.67	92.53
VGG	99.98	93.93
ResNet	99.99	95.35
DenseNet	99.99	95.74
DLA	99.99	95.93
Subnet1	99.98	92.74
Subnet2	99.92	92.17
Subnet3	99.93	92.70
Subnet4	99.71	91.34
Subnet5	99.76	92.44

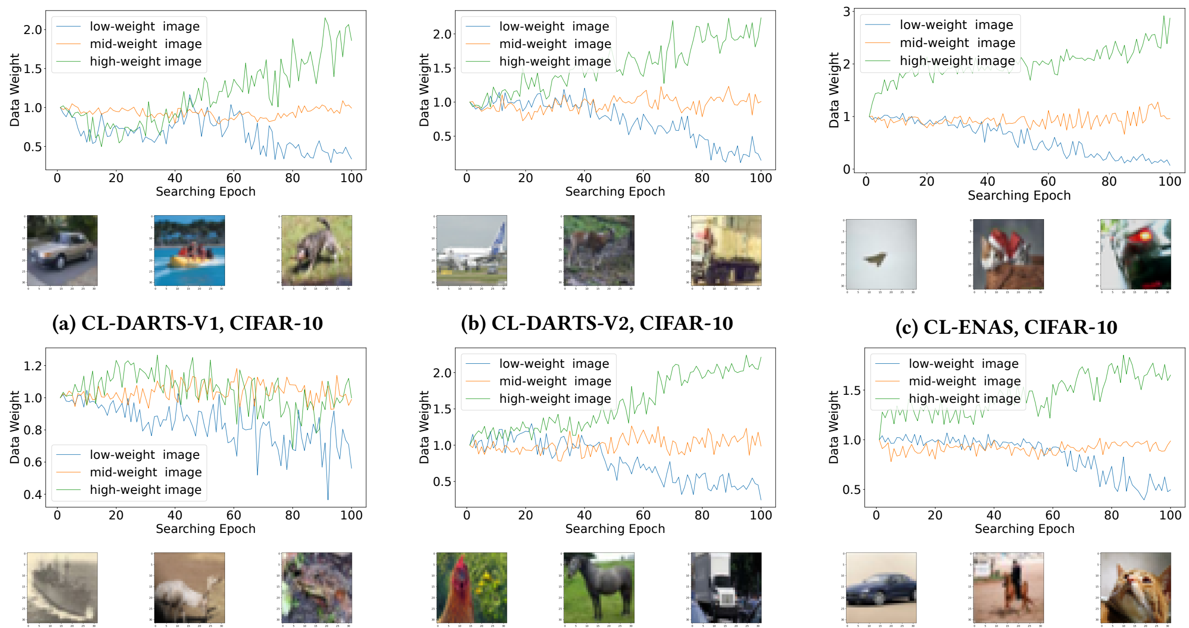
Our assumption is as follows. If all networks misclassify the same images, it demonstrates that some images are really difficult to learn. However, the fact is that the similarity scores of misclassification between arbitrary two networks are all below 0.01 on the training set and below 0.4 on the test set (, as is illustrated in Figure 1 of the main paper). Therefore, we conclude that different architectures fit different data samples and correspondingly different data has different influence on architectures.

2 CHANGING OF DATA WEIGHTS

In this section, we trace the weights changing of some typical data to analyze the effect of data uncertainty and prove that our method conducts weight-sharing NAS based on reweighted data instead of treating them equally.

As is illustrated in Fig 6, different data has different weight values and weight curves in the whole searching process of our Curriculum-NAS. Besides, it seems that the less uncertain image with low weight is generally more recognizable than the more uncertain one with high weight from the perspective of human. Therefore, the weights that we assign to data samples based on their uncertainty are meaningful and make sense.

[1] https://github.com/kuangliu/pytorch-cifar



(a) CL-DARTS-V1, CIFAR-10

(b) CL-DARTS-V2, CIFAR-10

(c) CL-ENAS, CIFAR-10

(d) CL-RSPS, CIFAR-10

(e) CL-GDAS, CIFAR-10

(f) CL-SETN, CIFAR-10

Figure 6: The changing of data weights in Curriculum-NAS. Each subfigure traces the weights of three images in CIFAR-10, including *low-weight image*: left image and blue line, *middle-weight image*: middle image and orange line, *high-weight image*: right image and green line.

REFERENCES

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

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

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

[4] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.

[5] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[6] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. 2018. Deep layer aggregation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2403–2412.