Supplementary Material for "Probeable DARTS with Application to Computational Pathology"

https://github.com/mahdihosseini/DARTS-ADP

1. Network Structures

The macro network structures in both the searching and evaluation phases are formed by stacking the normal and reduction cells sequentially. At $1/3$ and $2/3$ of the total depth of the network, there are reduction cells. Fig. 1 shows the general network structure, where the stem block contains several convolutional layers and the classifier consists of a global pooling layer and a fully connected layer.

The final architecture searched on ADP [3] is shown in Fig. 2. Note that there are no normal cells between the two reduction cells since the total number of cells is four, which is not divisible by three.

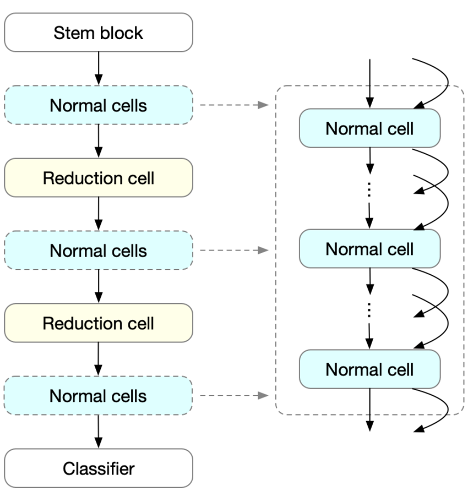

Figure 1. General network structure for searching and evaluation.

2. Dataset Details

CIFAR [6]. In the searching phase, we follow [8] to split the original training set into two parts, one for training and one for evaluation. In the evaluation phase, we use the default splits. We use random cropping with size 32x32 and random horizontal flipping as data augmentations.

CPath datasets. ADP and BCSS [1] are multi-label datasets, while BACH [2] and Osteosarcoma [7] are single-label. Their image resolution is all 272x272. We only conduct searching on ADP but evaluate the searched architecture on all four datasets. During searching, we treat half of the training set of ADP as the validation set. Data augmentations in all datasets include random horizontal and vertical flipping, random affine, and resize. Note that during searching on ADP, we resize the images to 64x64 to alleviate the computation overhead, and during evaluation, images are resized only in the test of different resolutions (136, 68, and 34).

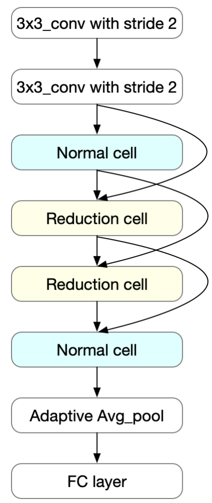

Figure 2. Final network structure searched on ADP.

3. Hyperparameters

3.1. Architecture Search

In CIFAR experiments, we train the network for 50 epochs with batch size 64 and initial channels 16. We test two optimizers for optimizing model weights, which are the original SGD [8] and Adas [4]. For DARTS+SGD, we follow [8] to use initial learning rate 0.025, cosine annealing scheduler, momentum 0.9 and weight decay 3×10^{-4}. For DARTS+Adas, we use initial learning rate 0.175, scheduler beta 0.98, momentum 0.9, and weight decay 3×10^{-4}. As for architecture parameter optimization, we follow [8] to use Adam [5] optimizer with initial learning rate 3×10^{-4}, momentum $(0.5, 0.999)$, and weight decay 10^{-3}.

In ADP experiments, most hyperparameters are the same except that we use batch size 32 due to computation overhead. We also increase the initial learning rate of DARTS+SGD to 0.175 for model weights optimization.

3.2. Architecture Evaluation

In both CIFAR and CPath experiments, we follow [8] to train the network for 600 epochs with batch size 96 and initial channels 36. We use SGD optimizer with an initial learning rate of 0.025, cosine annealing scheduler, momentum 0.9, and weight decay 3×10^{-4}. Additional enhancements include cutout and auxiliary towers as in [8]. Note that we disable auxiliary towers in training when we compare the performance of the searched architectures with the state-of-the-art networks.

References

[1] Mohamed Amgad, Habiba Elfandy, Hagar Hussein, Lamees A Atteya, Mai A T Elsebaie, Lamia S Abo Elnasr, Rokia A Sakr, Hazem S E Salem, Ahmed F Ismail, Anas M Saad, Joumana Ahmed, Maha A T Elsebaie, Mustafijur Rahman, Inas A Ruhban, Nada M Elgazar, Yahya Alagha, Mohamed H Osman, Ahmed M Alhusseiny, Mariam M Khalaf, Abo-Alela F Younes, Ali Abdulkarim, Duaa M Younes, Ahmed M Gadallah, Ahmad M Elkashash, Salma Y Fala, Basma M Zaki, Jonathan Beezley, Deepak R Chittajallu, David Manthey, David A Gutman, and Lee A D Cooper. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*, 35(18):3461–3467, 02 2019. 1

[2] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019. 1

[3] Mahdi S Hosseini, Lyndon Chan, Gabriel Tse, Michael Tang, Jun Deng, Sajad Norouzi, Corwyn Rowsell, Konstantinos N Plataniotis, and Savvas Damaskinos. Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11747–11756, 2019. 1

[4] Mahdi S Hosseini and Konstantinos N Plataniotis. Adas: Adaptive scheduling of stochastic gradients. *arXiv preprint arXiv:2006.06587*, 2020. 1

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[6] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[7] P Leavey, A Sengupta, D Rakheja, O Daescu, HB Arunachalam, and R Mishra. Osteosarcoma data from ut southwestern/ut dallas for viable and necrotic tumor assessment [data set]. *The Cancer Imaging Archive*, 14, 2019. `https://doi.org/10.7937/tcia.2019.bvhjhdas`. 1

[8] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018. 1, 2