

Efficient Neural Architecture Search on Low-Dimensional Data for OCT Image Segmentation

Nils Gessert¹

NILS.GESSERT@TUHH.DE

Alexander Schlaefer¹

SCHLAEFER@TUHH.DE

¹*Institute of Medical Technology, Hamburg University of Technology, Germany*

Abstract

Typically, deep learning architectures are handcrafted for their respective learning problem. As an alternative, neural architecture search (NAS) has been proposed where the architecture’s structure is learned in an additional optimization step. For the medical imaging domain, this approach is very promising as there are diverse problems and imaging modalities that require architecture design. However, NAS is very time-consuming and medical learning problems often involve high-dimensional data with high computational requirements. We propose an efficient approach for NAS in the context of medical, image-based deep learning problems by searching for architectures on low-dimensional data which are subsequently transferred to high-dimensional data. For OCT-based layer segmentation, we demonstrate that a search on 1D data reduces search time by 87.5 % compared to a search on 2D data while the final 2D models achieve similar performance.

Keywords: Neural Architecture Search, Deep Learning, Segmentation, OCT

1. Introduction

Over the last years, manual feature engineering has been replaced by deep learning approaches such as convolutional neural networks (CNNs) for numerous medical, image-based learning problems (Litjens et al., 2017). CNNs itself are often difficult to design and it is unclear what kind of architecture is suitable for which learning problem. Therefore, *neural architecture search* (NAS) has been proposed. Typical NAS approaches include grid search, genetic algorithms, bayesian optimization or random search (Kandasamy et al., 2018). Recently, reinforcement learning (RL) methods have been proposed where a recurrent controller is trained to predict an architecture’s structure by maximizing the architecture’s expected validation performance as a reward (Zoph and Le, 2016). This approach has been successful for 2D image classification problems (Liu et al., 2018; Zoph et al., 2018).

The concept of NAS is also very promising for the medical image domain as there is a vast amount of imaging modalities and learning problems that require architecture design. However, NAS can be very time-consuming which is even more problematic for medical image data which is often 3D or 4D in nature (Li et al., 2008). Some approaches have used lower dimensional data representations such as 2D slices instead of full 3D volumes in order to reduce computational effort (Litjens et al., 2017). However, many approaches have shown that considering higher dimensional context can improve performance (Kamnitsas et al., 2017; Gessert et al., 2018a,b).

We propose an efficient NAS approach for segmentation with mutlidimensional medical image data. To overcome long architecture search times, we perform the search on lower dimensional data which leads to shorter search times. Then, we transfer the learned architecture to the higher, target dimension. We show the concept for the example task of retinal layer segmentation with optical coherence tomography (OCT) data as the problem can be addressed in 1D (A-Scan segmentation) and 2D (B-Scan segmentation). Adopting the efficient neural architecture search (ENAS) framework (Pham et al., 2018), we learn submodules for a U-Net-like (Ronneberger et al., 2015) architecture. We demonstrate that our learned architecture outperforms a ResNet-inspired (He et al., 2016) baseline and that an architecture learned on 1D data transfers well to 2D data.

2. Methods

Dataset. We use a publicly available OCT dataset with images from patients with mild age-related macular degeneration (AMD) and normal subjects (Farsiu et al., 2014). Experts provided layer boundaries for the inner limiting membrane (ILM), retinal pigment epithelium drusen complex (RPEDC) and Bruchs membrane (BM). We generate pixel-wise annotations by assigning classes to tissue layers in between boundaries, i.e., ILM to RPEDC is class 1, RPEDC to BM is class 2 and BM to the end is class 3. The image space above the ILM is treated as background. Note that directly learning the boundaries can be beneficial for this problem (Roy et al., 2017). We chose a pixel-wise encoding to have a representative medical segmentation task that can be addressed with a standard U-Net.

Baseline Model. As a baseline we use a U-Net-like model. The model takes a 1D A-Scan or a 2D B-Scan as its input and predicts a segmentation map with the same size as the input. For the long-range connections we use summation, following (Yu et al., 2017). We use ResNet blocks in the network. Convolutions use a kernel size of 3 and extensions from 1D to 2D are performed by extending all kernels isotropically by an additional dimension.

ENAS U-Net. Next, we adopt the ENAS framework (Pham et al., 2018) for image classification to image segmentation with a U-Net. To simplify the architecture search space, we keep the general U-Net structure fixed and only learn new module blocks, similar to the micro search space in ENAS. The input/output and downsampling/upsampling layers also stay fixed. For the module search space, we let the controller learn the properties of 2 cells each containing 2 subcells. The cells’ output is the summation of the subcells’ output. For each subcell, the controller defines its input (the module input or another cell’s output) and its operation. Similar to ENAS, we allow five basic operations for the controller to choose from: convolutions with kernel size 3 or 5, average- and max-pooling with kernel size 3 and the identity transform.

Training and Evaluation. We consider a training set of 150 volumes (model training), a reward set of 56 volumes (controller training), a validation set of 2 volumes and a test set of 60 volumes. We follow ENAS with interleaved training of the model (dice loss) and the controller (dice score reward). After training for 200 epochs, we sample 20 architecture configurations from the controller and evaluate them on the validation set. Then, we select the best-performing configuration and retrain the model from scratch on the training set. Finally, we evaluate the model’s performance on the test set. For the baseline model, we train on the training set for 200 epochs and evaluate on the test set afterwards.

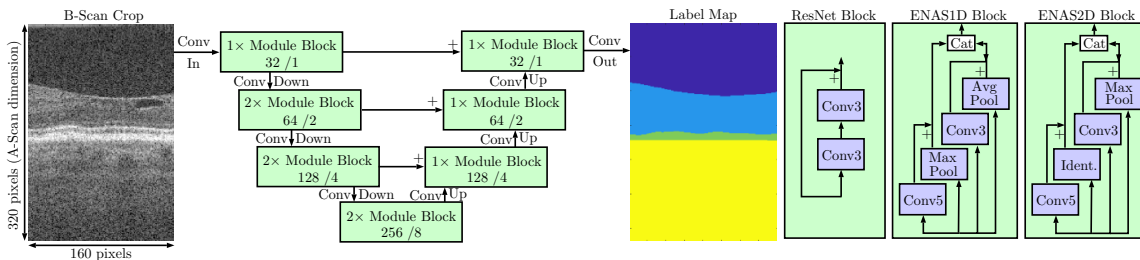


Figure 1: Left, the architecture with a 2D B-Scan input is shown. In each block the number of feature maps and total stride is given. Right, the baseline ResNet block and two learned ENAS blocks are shown.

Table 1: Dice and search time, performed on an NVIDIA GTX 1080 Ti.

Model	Mean Dice	Search Time
ResNet U-Net 1D	0.916 \pm 0.10	–
ENAS1D U-Net 1D	0.928 \pm 0.09	1 h 30 min
ResNet U-Net 2D	0.962 \pm 0.05	–
ENAS1D U-Net 2D	0.971 \pm 0.04	1 h 30 min
ENAS2D U-Net 2D	0.973 \pm 0.04	12 h 0 min

3. Results and Discussion

The architecture and the learned modules are shown in Figure 1. The results are shown in Table 1. Both the 1D and 2D architectures learned with ENAS on 1D data outperform the ResNet baseline. Notably, the increase is achieved without altering fundamental and potentially more impactful U-Net properties such as the encoder-decoder structure or the long-range connections. As a next step, these properties could be included in the search space which was successful for segmentation in the natural image domain with DeepLab-based architectures (Liu et al., 2019).

Performing a search on 1D data substantially decreases the search time by 87.5% compared to a search on 2D data while performance differences are marginal. This is particularly interesting as the OCT data is not isotropic and the spatial dimensions are quite different. This indicates that learning on low-dimensional, less resource demanding data representations is a viable approach for NAS. Thus, extension to other problems such as brain segmentation might be feasible, e.g., by performing NAS on axial slices before applying the discovered architectures on 3D volume data.

Summarized, we propose an efficient approach for NAS in the context of multidimensional medical image data. We demonstrate that searching for an architecture on low-dimensional data transfers well to high-dimensional data. An architecture discovered on 1D data performs similar to one discovered on 2D data while substantially reducing search time. Our approach could enable efficient NAS for a variety of medical learning problems.

Acknowledgments

This work was partially funded by the TUHH I^3 -Labs initiative.

References

- Sina Farsiu, Stephanie J Chiu, Rachelle V O’Connell, Francisco A Folgar, Eric Yuan, Joseph A Izatt, Cynthia A Toth, Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, et al. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*, 121(1):162–172, 2014.
- Nils Gessert, Jens Beringhoff, Christoph Otte, and Alexander Schlaefer. Force estimation from oct volumes using 3d cnns. *International journal of computer assisted radiology and surgery*, 13(7):1073–1082, 2018a.
- Nils Gessert, Matthias Schlüter, and Alexander Schlaefer. A deep learning approach for pose estimation from volumetric oct data. *Medical image analysis*, 46:162–179, 2018b.
- 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.
- Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.
- Kirthevasan Kandasamy, Willie Neiswanger, Jeff Schneider, Barnabas Poczos, and Eric P Xing. Neural architecture search with bayesian optimisation and optimal transport. In *Advances in Neural Information Processing Systems*, pages 2016–2025, 2018.
- Guang Li, Deborah Citrin, Kevin Camphausen, Boris Mueller, Chandra Burman, Borys Mychalczak, Robert W Miller, and Yulin Song. Advances in 4d medical imaging and 4d radiation therapy. *Technology in Cancer Research & Treatment*, 7(1):67–81, 2008.
- Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafourian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- 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.
- Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan Yuille, and Li Fei-Fei. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. *arXiv preprint arXiv:1901.02985*, 2019.

- Hieu Pham, Melody Y. Guan, Barret Zoph, Quoc V. Le, and Jeff Dean. Efficient neural architecture search via parameter sharing. In *ICML*, 2018.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Kattouzian, Christian Wachinger, and Nassir Navab. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, 8(8):3627–3642, 2017.
- Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *Thirty-first AAAI conference on artificial intelligence*, 2017.
- Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.
- 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.