

Invited:Hardware-aware Real-time Myocardial Segmentation Quality Control in Contrast Echocardiography

Dewen Zeng¹ Yukun Ding¹ Haiyun Yuan² Meiping Huang³ Xiaowei Xu³

Jian Zhuang³ Jingtong Hu⁴ Yiyu Shi¹

¹ Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

² Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

³ Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangzhou, China

⁴ Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA

Abstract—Automatic myocardial segmentation of contrast echocardiography has shown great potential in the quantification of myocardial perfusion parameters. Segmentation quality control is an important step to ensure the accuracy of segmentation results for quality research as well as its clinical application. Usually, the segmentation quality control happens after the data acquisition. At the data acquisition time, the operator could not know the quality of the segmentation results. On-the-fly segmentation quality control could help the operator to adjust the ultrasound probe or retake data if the quality is unsatisfied, which can greatly reduce the effort of time-consuming manual correction. However, it is infeasible to deploy state-of-the-art DNN-based models because the segmentation module and quality control module must fit in the limited hardware resource on the ultrasound machine while satisfying strict latency constraints. In this paper, we propose a hardware-aware neural architecture search framework for automatic myocardial segmentation and quality control of contrast echocardiography. We explicitly incorporate the hardware latency as a regularization term into the loss function during training. The proposed method searches the best neural network architecture for the segmentation module and quality prediction module with strict latency.

Index Terms—Neural Architecture Search, Image Segmentation, Quality Control, Contrast Echocardiography

I. INTRODUCTION

Deep learning models have shown remarkable success in automatic myocardial and left ventricle segmentation for contrast echocardiography [1]–[3]. After the images are segmented by deep neural networks (DNNs), the segmentation results can be used for further medical analysis such as quantifying myocardial perfusion parameters [2] and left ventricle ejection fraction [1]. However, even the state-of-the-art DNNs will occasionally fail due to low image quality or unexpected features of new data. In order to ensure accurate medical analysis, segmentation quality control needs to be employed to assess the segmentation quality and detect failure cases. Traditionally, both the segmentation module and the quality control module are applied after the data acquisition [1], [4], [5]. Even though such a framework can filter out good images for further analysis. One underlying problem is that once all the acquired images of a patient are low quality, the whole framework would not work. Therefore, it is desirable to design an on-the-fly segmentation and quality control framework to help the operator adjust the ultrasound probe or retake data if the image quality is unsatisfied during data acquisition.

However, building such a framework is non-trivial. It requires deploying the segmentation model and quality control model on the ultrasound machine with limited hardware resources, while satisfying the real-time requirement. Existing segmentation works mostly focus on handcrafted neural network architectures to improve the segmentation accuracy [2], [6]. In the meantime, some segmentation quality control works try to add additional cues such as shape and appearance

feature [7] or uncertainty [8], [9] to guide quality prediction. All of these works are computationally expensive that cannot achieve the latency requirement ($<50\text{ms}$) to avoid noticeable visual lag [10]. To enable real-time image segmentation and quality control, there are some works trying to solve one of them. For example, [11] proposed to model the 3D MRI videos as multiscale canonical form distributions and use a parallel statistical U-net to efficiently process these distributions and boost the inference speed. [12] used a light-weight CNN to predict the Dice of cardiovascular MR segmentation. The problem is that these methods are designed either from the segmentation perspective or from the quality control perspective, which may not be the optimal solution when combined together because the performance of these two modules are highly correlated. In this work, we argue that by jointly optimizing these two modules, we can achieve both high accuracy and low latency for the entire framework.

Neural architecture search (NAS) is a promising technique to achieve this. Recently, many works have empirically shown that NAS can jointly identify the best architecture and hardware design to maximize both accuracy and hardware efficiency [13]–[15]. To achieve hardware awareness, the overall latency is estimated on the target platform and jointly optimized with the network accuracy using reinforcement learning (RL) [13] or differentiable NAS [14]. In this work, we show that by applying the modern hardware-aware NAS technique to the problem of myocardial segmentation and quality control in contrast echocardiography, we can identify the best neural architecture for both segmentation module and quality control module with optimal accuracy and latency performance. Specifically, we design a supernet that includes a U-net like segmentation network and a Resnet like quality control network. The segmentation model outputs a segmentation mask and an uncertainty map, which will then be fit into the quality control model to predict the Dice. We use differentiable NAS [14] to simultaneously optimize the segmentation loss and quality control loss. In order to estimate the overall latency of the framework, we profile the latency of each operator used in the search space into a lookup table and model the network latency as a continuous function of the architecture parameters as in [14].

The main contributions of this paper are as follows:

- We build a neural architecture search framework that can identify the optimal segmentation network and quality control network for contrast echocardiography simultaneously.
- We explicitly incorporate the network latency on CPU as a regularization term into the loss function and jointly optimize the segmentation accuracy, quality prediction accuracy and hardware latency during searching.
- Experiment results on our collected Myocardial Contrast Echocardiography (MCE) dataset show that comparing to the handcrafted networks, the searched network can achieve 40.2% latency

reduction while maintaining comparable segmentation and quality prediction accuracy, suggesting that it can be used for real-time myocardial segmentation quality control of contrast echocardiography in clinical routine.

II. RELATED WORK

Segmentation Quality Control. Segmentation quality control is a technique to predict the segmentation quality of a given segmentation model (not just the original image) and to detect failure cases that are not suitable for further analysis. While some works predict segmentation quality using mostly shape and appearance features [7], DNNs have become popular in segmentation quality prediction along with their adoption for segmentation [12], [16]. Different uncertainty sources and different ways to aggregate pixel/voxel-level uncertainty to segment-level uncertainty are also explored [8], [9]. Other notable methods include Monte Carlo dropout [17] and Reverse Classification Accuracy (RCA) [18] but they are not suitable for real-time applications due to the high computational cost.

Neural Architecture Search. neural architecture search (NAS) is a technique that automatically searches for optimal neural architectures to replace the human effort in designing handcrafted neural architectures. In existing works, there are mainly three directions in searching an architecture: search space, search strategy, and performance estimation strategy [19]. The search space defines which architectures can be discovered by the algorithm. The search strategy mainly includes reinforcement learning [13], [20], evolutionary algorithms [21] and gradient-based methods [15], [22]. Performance estimation refers to the process of estimating the predictive performance of a candidate architecture on the target dataset. Recently, gradient-based differentiable NAS has been prevailing because of its high training speed and less hardware requirement [22]. Differentiable NAS models the searching problem as a single training process of an over-parameterized supernet that consists of multiple candidate paths. Each path is associated with some real-valued architecture parameters that are differentiable to the loss function. Therefore, standard gradient descent can be used to jointly train the weight parameters and architecture parameters. Recently, NAS-Unet [23], which uses a cell-based differentiable NAS strategy, shows great success in searching the best neural architecture for medical image segmentation. However, hardware latency is not considered in their method, which hinders its utilization in real-time situations.

III. METHOD

In this section, we will first introduce the framework overview in Section III-A. Then, we will discuss the search space of the segmentation network and quality control network in Section III-B and Section III-C. Finally, The hardware-aware loss function will be discussed in Section III-D.

A. Framework Overview

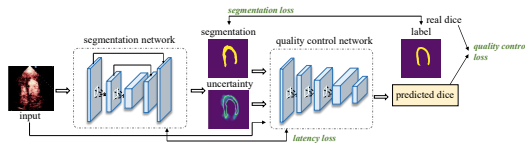


Fig. 1: Overview of the proposed NAS framework. The framework jointly searches for the optimal segmentation network and quality control network. In the meantime, the overall hardware latency of the framework is estimated and used as an additional term to the final loss to penalize the network's latency on the target device (CPU).

Fig. 1 shows the overview of our proposed NAS framework. The framework consists of a U-net like segmentation network as well as a Resnet like quality control network that needs to be searched. The segmentation network takes RGB contrast echocardiography as the input and outputs the segmentation result and an uncertainty map. Then the segmentation result and the uncertainty map, along with the original image are concatenated and fit into the quality control model to predict the estimated Dice. The ground truth of Dice is computed by using the segmentation result and ground truth mask of myocardial. The segmentation loss (i.e. cross-entropy loss), quality control loss (i.e. mean square loss), and the estimated latency of the whole framework will be used in the final loss function during searching. As such, the segmentation network and quality control network can be optimized and searched simultaneously with better accuracy and latency.

B. Segmentation Network Search Space

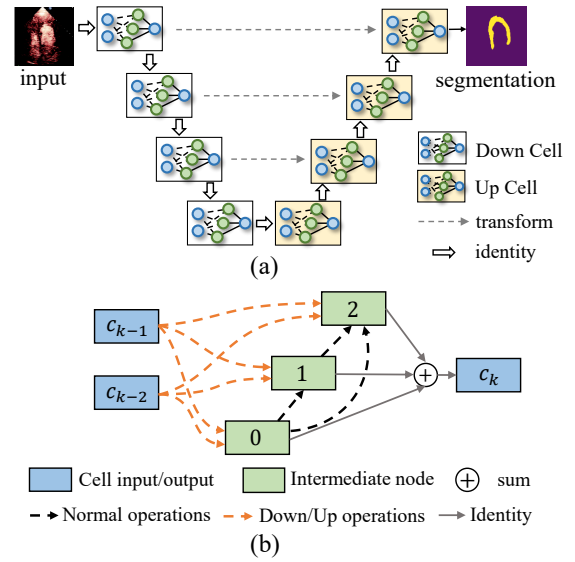


Fig. 2: (a) The U-net like segmentation supernet. The network stacks two kinds of cell structures: Down cell and Up cell, which need to be searched. The gray arrow represents the transform operation that belongs to the Up cell. (b) Detail cell structure. Each dash arrow represents the operation that needs to be searched from a predefined primitive operation pool. Down and Up operations are used in Down Cell and Up cell respectively.

To search for the optimal segmentation network, we use the search space like NAS-Unet [23], which has been proved to be useful for several medical image datasets. The overall segmentation supernet is shown in Fig. 2(a). Two kinds of cell structures called Down cell and Up cell are defined and stacked to form a U-net like backbone. They share the same network structure as shown in Fig. 2(b). The inputs of the cell are defined as the cell out of the previous two cells [22]. The dash edges/lines represent the operations (e.g., convolution operation and pooling operation) that need to be searched between two feature maps (either cell input or intermediate node). Suppose there are m intermediate nodes inside a cell, the total number of operations need to search is $2m + m(m+1)/2$. The output of a cell is defined as the sum of all intermediate nodes.

Table I shows the three types of primitive operation sets used for searching cells. Following the work of [22], we use 3×3 normal convolution, dilated convolution, and separable convolution. The down convolution is the stride 2 version of normal convolution, and

TABLE I: Three types of primitive operation set. For all convolution operations and pooling operations, we use kernel size 3×3 .

Down operations	Up operation	Normal operation
down conv	up conv	conv
down dilated conv	up dilated conv	dilated conv
down separable conv	up separable conv	separable conv
max pooling	-	identity
average pooling	-	zero

the up convolution represents the 2d transpose convolution. Pooling operations are only used in down operation set. Identity and zero operations are only used in normal operation set.

A cell can be viewed as a directed acyclic graph (DAG) with N nodes (including the input and output). Since each intermediate node x_i represents a feature map and each edge (i, j) is associated with a primitive operation $op^{(i,j)}$, each intermediate node x_j inside a cell can be computed by all of its previous nodes:

$$x_j = \sum_{i < j} \sum_{k \in K} p_{\alpha_k^{(i,j)}} \cdot op_k^{(i,j)}(x_i), j > 1, \quad (1)$$

$$p_{\alpha_k^{(i,j)}} = \frac{\exp(\alpha_k^{(i,j)})}{\sum_{k \in K} \exp(\alpha_k^{(i,j)})}. \quad (2)$$

where $op_k^{(i,j)}$ is the k -th primitive operation from K candidates in the operation set. $\alpha_k^{(i,j)}$ is the weight of the corresponding primitive operation k on the edge (i, j) , and $p_{\alpha_k^{(i,j)}}$ is the probability of choosing this operation among K candidates. After neural architecture search, we can prune the redundant operations and paths by choosing $\alpha_k^{(i,j)}$ with the highest probability.

C. Quality Control Network Search Space

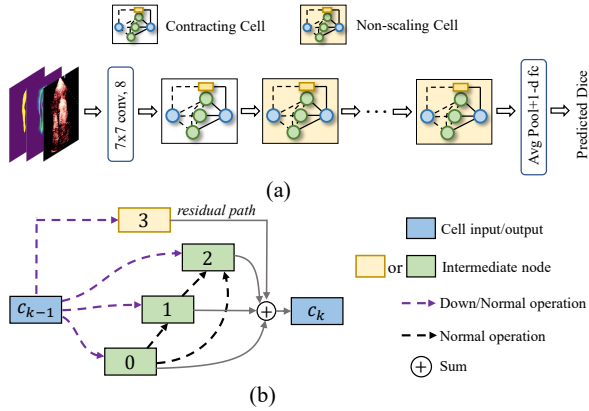


Fig. 3: (a) The Resnet like quality control model. The network consists of two kinds of cell structures, namely Contracting Cell and Non-scaling Cell. Two types of cells are stacked in an alternating way. (b) Detail cell structure. Dashed lines/edges represent the operations need to search. The purple edges denote the Down operation and Normal operation in the Contracting cell and Non-scaling cell respectively.

Our quality control NAS network is shown in Fig. 3(a). In this network, we define two types of cell structures, namely Contracting cell and Non-scaling cell to mimic the resnet backbone, they share the same network structure. Fig. 3(b) demonstrate the detail structure of the cells. Unlike the cell structure in the segmentation network, the input of Contracting/Non-scaling cell is the cell output of the

previous layer. Inside a cell, a residual path (see Fig. 3(b)) is used to mimic the residual behavior. The down and normal operations, whose candidates are the same as in Section III-B, are used in Contracting cell and Non-scaling cell respectively. The computation of feature map in the cell is also the same as in Section III-B.

D. Hardware-aware Loss Function

In order to jointly optimize the segmentation accuracy, quality prediction accuracy and hardware latency, the following loss function is used during searching:

$$\min : \mathcal{L}(w, \alpha) = \mathcal{L}_{CE}(w, \alpha) + \lambda_1 \mathcal{L}_{MSE}(w, \alpha) + \lambda_2 LAT(\alpha). \quad (3)$$

where w is the network weights and α is the architecture parameters. The $\mathcal{L}_{CE}(w, \alpha)$ is the cross-entropy loss for the segmentation network. $\mathcal{L}_{MSE}(w, \alpha)$ is the mean squared error loss for the quality prediction network. $LAT(\alpha)$ is the expected latency of the whole framework. λ_1 and λ_2 are two scaling factors that control the trade-off between segmentation accuracy, quality prediction accuracy and hardware latency. λ_1 and λ_2 are set to 1 and 0.001 in our experiments.

Similar to the latency estimation strategy in FBNet [14], we use a differentiable method to estimate the expected latency of the cell and then the latency of the framework will be the summation of all the stacked cells in the network. Specifically, the latency of each candidate primitive operation is profiled into a lookup table for efficiency estimation. During training, the estimated latency of each edge (i, j) that needs to search in a cell is computed by summing up the weight latency of all candidate operations.

$$latency(edge^{(i,j)}) = \sum_{k \in K} GS(\alpha_k^{(i,j)} | \alpha^{(i,j)}) latency(op_k^{(i,j)}). \quad (4)$$

in which $GS(\cdot)$ is the Gumbel-Softmax sampling rule to make the architecture parameters differentiable to the latency loss term. $latency(op_k^{(i,j)})$ is a constant coefficient from the lookup table. Then, the estimated latency of a cell is the summation of the estimated latency of all edges, and the estimated latency of whole network is the summation of estimated latency of all cells.

IV. EXPERIMENTS

A. Experimental Setup

Dataset and Evaluation Protocols. The dataset we use consists of 100 patients who received MCE from Guangdong Provincial People's Hospital. Each patient has an MCE sequence containing around 100 frames in A4C chamber view. We randomly sample 30 frames from the MCE sequence of each patient, which results in 3000 images in total in this dataset. All images are center cropped and resized to a fixed resolution 256×256 . During searching, we split the 100 patients into D_{train} and D_{test} by a ratio of 8:2. Since the quality control network needs to be trained on a dataset that is not used for training the segmentation network to avoid over-fitting, we further use 40 patients (denoted as D_{seg_train}) in D_{train} for training the segmentation network and use the rest 40 patients (denoted as D_{seg_eval}) for evaluating the segmentation network. The quality control network is trained on the whole D_{train} . When searching on D_{seg_eval} , both networks are optimized using the loss in Equation 3. When D_{seg_eval} is used, the segmentation network is frozen and only the parameters of the quality control network are updated. Similar to [22], we use the first-order approximation optimization strategy for searching, where the network weight parameters and architecture parameters are updated in turn. After searching, for both networks, we randomly sample 60 subjects for training, the rest are used for validation. The evaluation protocols include segmentation Dice, mean

absolute error (MAE) of the predicted Dice, and the overall framework latency on CPU. All the latency is emulated on a 4-Core Intel i7 processor which is used on the Philips Epiq 7 ultrasound machine.

Setup. In our NAS supernet, the intermediate nodes of all the cells are set to 3. The initial filter size is set to 4 and 8 for segmentation network and quality control network for latency reduction. Data augmentation such as translation, rotation and scale are used during searching. We search the network for 80 epochs in total, the first half is used to optimize the network weight. For the rest 40 epochs, network weights and architecture parameters are optimized alternately. We repeat the experiments five times with different random seeds and report the best validation performance of the architectures. All experiments run on a machine with four NVIDIA GeForce GTX 1080 GPUs.

B. Results

TABLE II: Comparison between baseline results and our proposed hardware-aware NAS on Dice, MAE and latency. U-net(8) means the vanilla U-net with initial filter size 8.

Method	Dice	MAE	Latency (ms)	FLOPs (M)
U-net(8)+Resnet18	0.797	0.050	47	1045
U-net(16)+Resnet18	0.839	0.049	82	3635
w/o hardware-aware	0.837	0.047	54	440
hardware-aware	0.832	0.049	49	373

Table II shows the comparison results among the baselines and the proposed hardware-aware NAS. It can be seen that compared to handcrafted neural architecture U-net(16)+Resnet18, the searched network reduces the inference latency by 40.2% with very small accuracy degradation. Integrating hardware latency during searching can further reduce the latency of the search architecture by 5ms. All searched cell structures are shown in Fig. 4.

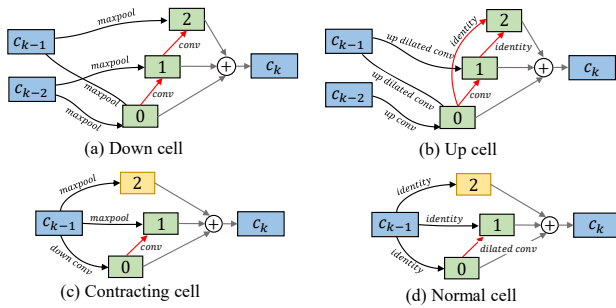


Fig. 4: Illustration of searched cell structures, including the Down Cell, Up cell in segmentation network and the Contracting cell and Non-scaling cell in quality control network.

V. CONCLUSION

In this paper, we proposed a hardware-aware NAS framework to jointly search for the optimal myocardial segmentation network and quality control network for contrast echocardiography with better accuracy and latency for the real-time requirement. Experimental results on our MCE dataset show that searched networks can greatly reduce the inference time of the entire framework for real-time ultrasound acquisition guidance, while still maintaining high accuracy.

REFERENCES

[1] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, "Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model," *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1081–1091, 2017.

[2] M. Li, D. Zeng *et al.*, "A deep learning approach with temporal consistency for automatic myocardial segmentation of quantitative myocardial contrast echocardiography," *The International Journal of Cardiovascular Imaging*, pp. 1–12, 2021.

[3] S. Dong *et al.*, "Voxelatlasgan: 3d left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 622–629.

[4] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Inherent brain segmentation quality control from fully convnet monte carlo sampling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 664–672.

[5] R. Robinson, V. V. Valindria, W. Bai, O. Oktay, B. Kainz, H. Suzuki, M. M. Sanghvi, N. Aung, J. M. Paiva, F. Zemrak *et al.*, "Automated quality control in image segmentation: application to the uk biobank cardiovascular magnetic resonance imaging study," *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1, pp. 1–14, 2019.

[6] L.-C. Chen, G. Papandreou *et al.*, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[7] B. Ruijsink *et al.*, "Fully automated, quality-controlled cardiac analysis from cmr: validation and large-scale application to characterize cardiac function," *Cardiovascular Imaging*, vol. 13, no. 3, pp. 684–695, 2020.

[8] K. Hoebe, V. Andrearczyk, A. Beers, J. Patel, K. Chang, A. Depeursinge, H. Müller, and J. Kalpathy-Cramer, "An exploration of uncertainty information for segmentation quality assessment," in *Medical Imaging 2020: Image Processing*, vol. 11313. International Society for Optics and Photonics, 2020, p. 113131K.

[9] A. Jungo, R. Meier, E. Ermis, E. Herrmann, and M. Reyes, "Uncertainty-driven sanity check: Application to postoperative brain tumor cavity segmentation," *arXiv preprint arXiv:1806.03106*, 2018.

[10] M. Annett, A. Ng, P. Dietz, W. F. Bischof, and A. Gupta, "How low should we go? understanding the perception of latency while inking," in *Proceedings of Graphics Interface 2014*, 2014, pp. 167–174.

[11] T. Wang, J. Xiong *et al.*, "Msu-net: Multiscale statistical u-net for real-time 3d cardiac mri video segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 614–622.

[12] R. Robinson *et al.*, "Real-time prediction of segmentation quality," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 578–585.

[13] W. Jiang *et al.*, "Accuracy vs. efficiency: Achieving both through fpga-implementation aware neural architecture search," in *Proceedings of the 56th Annual Design Automation Conference 2019*, 2019, pp. 1–6.

[14] B. Wu, X. Dai, P. Zhang *et al.*, "Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10734–10742.

[15] H. Cai *et al.*, "Proxlessnas: Direct neural architecture search on target task and hardware," *arXiv preprint arXiv:1812.00332*, 2018.

[16] R. Zhang and A. C. Chung, "A fine-grain error map prediction and segmentation quality assessment framework for whole-heart segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 550–558.

[17] A. G. Roy, S. Conjeti *et al.*, "Bayesian quicknat: model uncertainty in deep whole-brain segmentation for structure-wise quality control," *NeuroImage*, vol. 195, pp. 11–22, 2019.

[18] V. V. Valindria, I. Lavdas, W. Bai, K. Kamnitsas, E. O. Aboagye *et al.*, "Reverse classification accuracy: predicting segmentation performance in the absence of ground truth," *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1597–1606, 2017.

[19] T. Elsken, J. H. Metzen, F. Hutter *et al.*, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, no. 55, pp. 1–21, 2019.

[20] B. Baker *et al.*, "Designing neural network architectures using reinforcement learning," *arXiv preprint arXiv:1611.02167*, 2016.

[21] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu *et al.*, "Large-scale evolution of image classifiers," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2902–2911.

[22] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," *arXiv preprint arXiv:1806.09055*, 2018.

[23] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "Nas-unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44 247–44 257, 2019.