Medical Image Analysis 82 (2022) 102589

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Auto-DenseUNet: Searchable neural network architecture for mass segmentation in 3D automated breast ultrasound ☆

Xuyang Cao [a], Houjin Chen [a], Yanfeng Li [a,*], Yahui Peng [a], Yue Zhou [a], Lin Cheng [b], Tianming Liu [c], Dinggang Shen [d,e]

[a] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, China
[b] Peking University People's Hospital, Beijing, 100044, China
[c] Department of Computer Science, University of Georgia, Athens, GA 30602, USA
[d] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
[e] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

ARTICLE INFO

*Keywords:*
ABUS image
Breast mass segmentation
Neural architecture search
Deep learning

ABSTRACT

Accurate segmentation of breast mass in 3D automated breast ultrasound (ABUS) plays an important role in breast cancer analysis. Deep convolutional networks have become a promising approach in segmenting ABUS images. However, designing an effective network architecture is time-consuming, and highly relies on specialist's experience and prior knowledge. To address this issue, we introduce a searchable segmentation network (denoted as Auto-DenseUNet) based on the neural architecture search (NAS) to search the optimal architecture automatically for the ABUS mass segmentation task. Concretely, a novel search space is designed based on a densely connected structure to enhance the gradient and information flows throughout the network. Then, to encourage multiscale information fusion, a set of searchable multiscale aggregation nodes between the down-sampling and up-sampling parts of the network are further designed. Thus, all the operators within the dense connection structure or between any two aggregation nodes can be searched to find the optimal structure. Finally, a novel decoupled search training strategy during architecture search is also introduced to alleviate the memory limitation caused by continuous relaxation in NAS. The proposed Auto-DenseUNet method has been evaluated on our ABUS dataset with 170 volumes (from 107 patients), including 120 training volumes and 50 testing volumes split at patient level. Experimental results on testing volumes show that our searched architecture performed better than several human-designed segmentation models on the 3D ABUS mass segmentation task, indicating the effectiveness of our proposed method.

## 1. Introduction

Automated breast ultrasound (ABUS) is an emerging imaging modality, complementary with mammogram, for breast cancer diagnosis, as it can improve sensitivity of cancer detection in dense breasts (van Zelst and Mann, 2018; Vourtsis and Kachulis, 2018). Besides, compared with mammogram and 2D ultrasound, the 3D ABUS can show multi-slices and has higher reproducibility, which is important for temporal comparison (Kozegar et al., 2018). Segmentation of breast mass in ABUS is essential in breast cancer analysis, such as surgical planning, 3D lesion construction, and designing of computer-aided diagnosis (CAD) systems (Tan et al., 2012; Gómez-Flores and Ruiz-Ortega, 2016). However, mass segmentation in ABUS is challenging, due to serious artifacts in ABUS images, ambiguous mass boundaries, and large shape and size variations of breast masses, as shown in Fig. 1.

In recent years, due to powerful automated representation capability of convolutional neural networks (CNNs), deep learning based methods have made remarkable progress in medical image analysis community (Zhou et al., 2021b; Cao et al., 2021; Wang et al., 2019a; Kumar et al., 2020). It has been proven that the performance of CNNs and the representations of data are highly determined by network architecture (Ren et al., 2020). Therefore, a great number of works have been conducted to build more effective network architectures (Liu et al., 2020a; Wang et al., 2019b; Chu et al., 2020; Nie et al., 2019). However, the common drawback of these human-designed

☆ This work was supported by the National Natural Science Foundation of China (grant numbers 61872030, 62172029).
* Corresponding author.
*E-mail address:* yf.li@bjtu.edu.cn (Y. Li).

https://doi.org/10.1016/j.media.2022.102589
Received 8 March 2021; Received in revised form 18 July 2022; Accepted 17 August 2022
Available online 23 August 2022

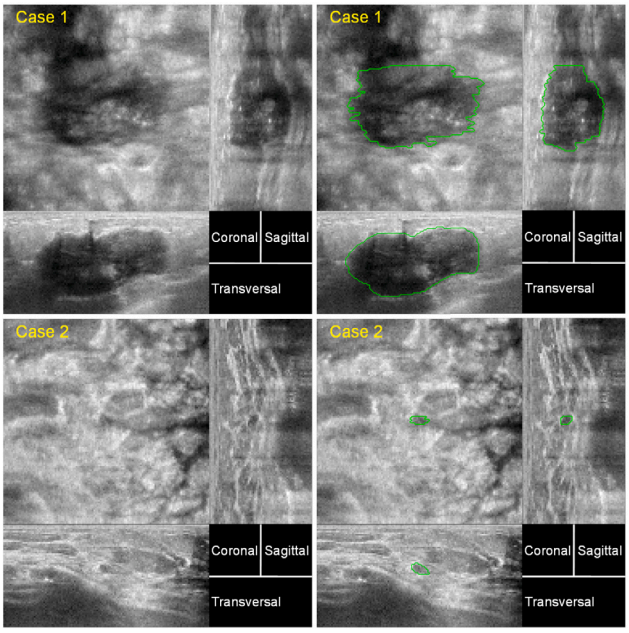


**Fig. 1.** Two breast mass samples/cases in our ABUS dataset. Note (1) serious artifacts and ambiguous boundaries in these samples, and (2) huge size variation across different samples in the dataset, i.e., the volume size is around 12,878 $mm^3$ for the first sample (Case 1) while only around 29 $mm^3$ for the second sample (Case 2).

architectures is that they rely highly on researchers' prior knowledge and experience. Besides, it is also difficult for researchers to think outside of the box in many cases. As a result, the designed architecture may not be the optimal, which negatively affects the performance of designed networks.

To solve the aforementioned problem, the neural architecture search (NAS) method has been proposed (Zoph and Le, 2017; Ren et al., 2020). The basic idea of NAS is to reduce the intervention of human and enable the network design architecture automatically for a specific task. It is worth noting that the recent NAS based methods have successfully outperformed human-designed architectures in both natural and medical image processing tasks (Liu et al., 2018; Zoph et al., 2018; Real et al., 2019; Ji et al., 2020; Li et al., 2020), indicating great potential of NAS. In this paper, we will mainly study the 3D ABUS mass segmentation problem based on the NAS approach.

Specifically, we develop a NAS-based segmentation framework (namely, Auto-DenseUNet) to reduce dependence on expert's domain knowledge when segmenting breast mass in ABUS images. To this end, we first design a multiscale segmentation search space that consists of three different kinds of cells, including reduction cell, normal cell, and expansion cell. Following the classic encoder–decoder structure, the reduction cell and expansion cell are designed to extract feature representations at different scales. In particular, the reduction cell has a densely connected structure, which encourages information and gradient flows throughout the whole network. The normal cell is used to aggregate feature maps at different scales to enrich semantic representation. By these newly designed cells, our search space is specifically suitable for masses with large size variation. To perform effective network search, we investigate the continuous relaxation (Liu et al., 2019b), which relaxes the discrete search space into the continuous space. In this way, the problem of identifying the optimal architecture is formulated as an optimization problem that can be solved by highly effective gradient descent methods. However, the continuous relaxation consumes too much computational memory in the searching process, especially in 3D medical image segmentation tasks. To solve this issue, we design a decoupled search training strategy in this paper. Specifically, we utilize two hyper-parameters to control the number of feature maps through the entire network. Then, we search the architecture with smaller number of feature maps to alleviate the memory limitation and retrain the searched network with larger number of feature maps to enrich feature representation. Comprehensive experimental results show obvious improvement of our proposed method in ABUS mass segmentation, compared with human-designed segmentation networks as well as existing ABUS mass segmentation methods.

The contributions of this paper are summarized as follows:

- We propose Auto-DenseUNet, a NAS-based mass segmentation framework with densely connected structure and multiscale information fusion scheme, for automatically searching the optimal segmentation architecture in a pre-defined search space.
- We introduce a decoupled search training strategy into the Auto-DenseUNet, for relieving the computing memory burden during architecture searching and encouraging the searched network to learn richer feature representation during the training.

## 2. Related works

### 2.1. ABUS mass segmentation

**Hand-crafted-features based Methods:** Early ABUS mass segmentation were mainly based on hand-crafted features (Kozegar et al., 2018; Agarwal et al., 2017; Tan et al., 2016; Araújo et al., 2017; Xi et al., 2017). For example, Agarwal et al. (2017) developed a semi-automatic framework for breast lesion segmentation based on the watershed algorithm. This framework achieved a mean Dice similarity coefficient (DSC) of 0.69 in 56 volumes. Tan et al. (2016) utilized a depth-guided dynamic programming based on spiral scanning method to segment malignant lesions in the 3D ABUS images, and reported a DSC of 0.73 on a dataset of 73 cancers. Xi et al. (2017) proposed a mass segmentation method with prior knowledge learning of abnormal mass regions. They first extracted a series of hand-crafted features on the super-pixels of ABUS images, and then applied a SVM classifier to classify these super-pixels into different classes. Finally, the learned prior knowledge was fine-tuned by a level-set segmentation algorithm. Kozegar et al. (2018) proposed a two-stage approach that incorporated shape information into the segmentation process. At the first stage, they utilized an adaptive region growing algorithm to generate a rough boundary of breast mass. In the second stage, they used the rough boundary as an initial contour for the level-set method (Li et al., 2010). But, the common limitation of all these traditional algorithms is that the segmentation performance is dependent on hand-crafted features.

**Deep Learning based Methods:** Recent progresses on ABUS mass segmentation are mainly from deep learning based methods (Xing et al., 2020; Cao et al., 2020; Zhou et al., 2021a; Wang et al., 2020). In most cases, these deep learning based methods perform better than the traditional hand-crafted-feature based algorithms. For example, based on the classic VNet (Milletari et al., 2016), Zhou et al. (2021a) developed a multi-task framework combining breast mass segmentation and classification into one network, to promote the two tasks each other. They further utilized an iterative feature-refining mechanism to suppress noise in ABUS images. To alleviate the class imbalance between foreground and background, Cao et al. (2020) introduced a combined region and boundary constrained loss function to improve the segmentation performance on an improved VNet. This method achieved a DSC of 0.82 on a dataset of 83 volumes. Xing et al. (2020) proposed a semi-pixel-wise cycle generative adversarial network for segmenting breast lesion in 2D ultrasound images. They reported a DSC improvement of 2% and 13% over the fully convolutional neural network and the level-set method, respectively. Wang et al. (2020) proposed a densely supervised network, utilizing the pre-trained C3D model (Tran et al., 2015), to alleviate the over-fitting problem. They further introduced a threshold map to supply adaptive thresholds for

classifying the probability maps. Cao et al. (2021) constructed a dilated densely connected UNet to address the large shape and size of breast masses in ABUS data, and reported better performance than several classic segmentation networks. Although these deep learning based methods benefit from automatic feature representations than the traditional segmentation methods, designing effective network architecture highly relies on expert's domain knowledge.

### 2.2. Feature aggregation

The skip connections introduced in the encoder–decoder networks (Ronneberger et al., 2015; Kamnitsas et al., 2017) can be seen as a simple case of feature aggregation. Although these networks have the ability to fuse structural information from encoder to decoder, there is no solid evidence that can prove they have optimal structures for feature fusion. Consequently, hierarchical feature aggregation scheme has been explored in recent literature (Zhao et al., 2020; Zhou et al., 2020; Tian et al., 2019; Ruan et al., 2020). The Triple-UNet (Zhao et al., 2020) introduced a progressive dense feature aggregation, where the network was encouraged to fuse features progressively from different branches. Zhou et al. (2020) proposed UNet++ through an ensemble of U-Nets of varying depths. They further utilized skip connections to aggregate features in different semantic scales for highly flexible feature fusion scheme. Tian et al. (2019) designed a data-dependent upsampling (DUpsampling) to replace bilinear interpolation in the decoder of the network. The DUpsampling has better flexibility in leveraging almost arbitrary combinations of features from the CNN encoder. The above operations in aggregation nodes and the connections between different nodes were designed by human expert. In this paper, we construct a searchable multiscale and multi-label aggregation scheme to automatically select the best operations and connections.

### 2.3. Efficient methodologies for image segmentation

Designing an effective deep learning based framework for a specific task requires high-level domain knowledge and experience of human designers. Especially for 3D medical image analysis, a small error may lead to large performance decrease (Litjens et al., 2017). Consequently, efficient methodologies that reduce the interaction of expert and automatically determine the configurations of deep learning approaches are desired. One solution to this problem is the rule-based methods (Barzegar and Jamzad, 2019; Isensee et al., 2021) that formulate the dependencies between specific dataset properties and design choices, such as nnUNet (Isensee et al., 2021). The rule-based parameters in nnUNet include the architecture topology, intensity normalization, resampling, and batch size, etc. The other solution to this problem is AutoML, which targets a progressive automation of machine learning, based on principles of optimization and machine learning itself (Hutter et al., 2019). Generally, AutoML includes hyper-parameter optimization (HPO) (Luo, 2016; Mackay et al., 2019), meta-learning (Hospedales et al., 2021) and NAS. The HPO aims to automatically set deep learning hyper-parameters to optimize performance. The meta-learning learns how to learn across tasks, while the NAS focuses on designing more effective network architectures for specific tasks. In the following, we mainly review the NAS-based approaches both in natural image segmentation and medical image segmentation.

**Natural Image Segmentation with NAS:** The effectiveness of NAS was mainly verified on classification task in previous studies (Zoph et al., 2018; Nekrasov et al., 2019), since classification is among the most fundamental tasks in vision recognition. Generally, applying NAS techniques from classification to segmentation tasks have the following requirements: (1) more relaxed and general search space to accommodate higher image resolution, and (2) more efficient search method to meet heavier computation caused by higher resolution. Therefore, Liu et al. (2019a) designed a hierarchical architecture search space to search the architecture both in cell and network levels. The cell architecture is defined as small repeated modules that extract feature representations, while network architecture is used to control the optimal spatial change. Chen et al. (2020) proposed latency regularization to make a balance between high accuracy and low latency. In the spirit of knowledge distillation (Hinton et al., 2015), they further designed a new co-searching framework, where the performance is boosted by the teacher–student distillation. Although above methods achieved satisfactory performance for natural image segmentation, directly applying these models to medical image segmentation may suffer from performance degradation, due to obvious difference between natural and medical images. For example, the 3D ABUS images used in this paper require much higher computational resources than the 2D natural images, and the signal to noise ratio (SNR) of 3D ABUS images is also much smaller than most 2D natural images.

**Medical Image Segmentation with NAS:** The study of NAS-based medical image segmentation is at the early stage. The SCNAS (Kim et al., 2019) constructed an encoder–decoder search space, which included five different types of cells. Each cell was represented as a directed acyclic graph (DAG), and operators and connections could be searched inside the DAG. Different from the SCNAS that utilized differentiable NAS, the RONASMIS model (Bae et al., 2019) introduced resource optimized discrete NAS scheme on 3D medical image segmentation, and achieved better performance than the SCNAS on several datasets. Although attention mechanism has achieved satisfying results in medical image segmentation, the usage of spatial or channel attention module is mainly determined by designer's experience. Consequently, Liu et al. (2020b) designed a search space to search the optimal attention modules. Besides, a synchronous search approach was utilized to search architectures independently for different attention modules. Ji et al. (2020) introduced an UXNet to search multi-level feature aggregation for 3D medical image segmentation. Similar to Ji et al., in this paper, we also utilize the NAS to search multiscale feature fusion scheme. However, there are several differences between Ji et al.'s (Ji et al., 2020) and ours. First, our search space has higher degree of freedom and richer feature representations than those in Ji et al. (2020). We use densely-connected structures in reduction cell to encourage richer feature representations, and each reduction cell has more paths to search than those in Ji et al. (2020). Besides, our method has better information and gradient flow due to the use of the densely-connected structures. Furthermore, with the decoupled search training strategy, our search space is deeper and more flexible for 3D medical image segmentation.

**Table 1**
Number of volumes in different types of breast masses and volume size groups in the ABUS dataset.

| | Types of masses | | Volume sizes of masses ($mm^3$) | | | |
|---|---|---|---|---|---|---|
| | Benign | Malignant | (0, 500] | (500, 1000] | (1000, 3000] | (3000, 20 000] |
| Total | 51 | 119 | 66 | 29 | 54 | 21 |
| Training | 39 | 81 | 53 | 18 | 37 | 12 |
| Testing | 12 | 38 | 13 | 11 | 17 | 9 |

## 3. Materials and methods

### 3.1. Materials

The ABUS dataset used in this paper was obtained from Peking University People's Hospital using the freehand ACUSON 2000 Ultrasound System (Siemens). Each patient has four volumes from different views, including left lateral (LLAT), left medial (LMED), right lateral (RLAT), and right medial (RMED). To conduct mass segmentation, we only select volumes with breast masses. A total of 170 volumes (from 107 patients) are used. To evaluate the proposed method, we randomly split

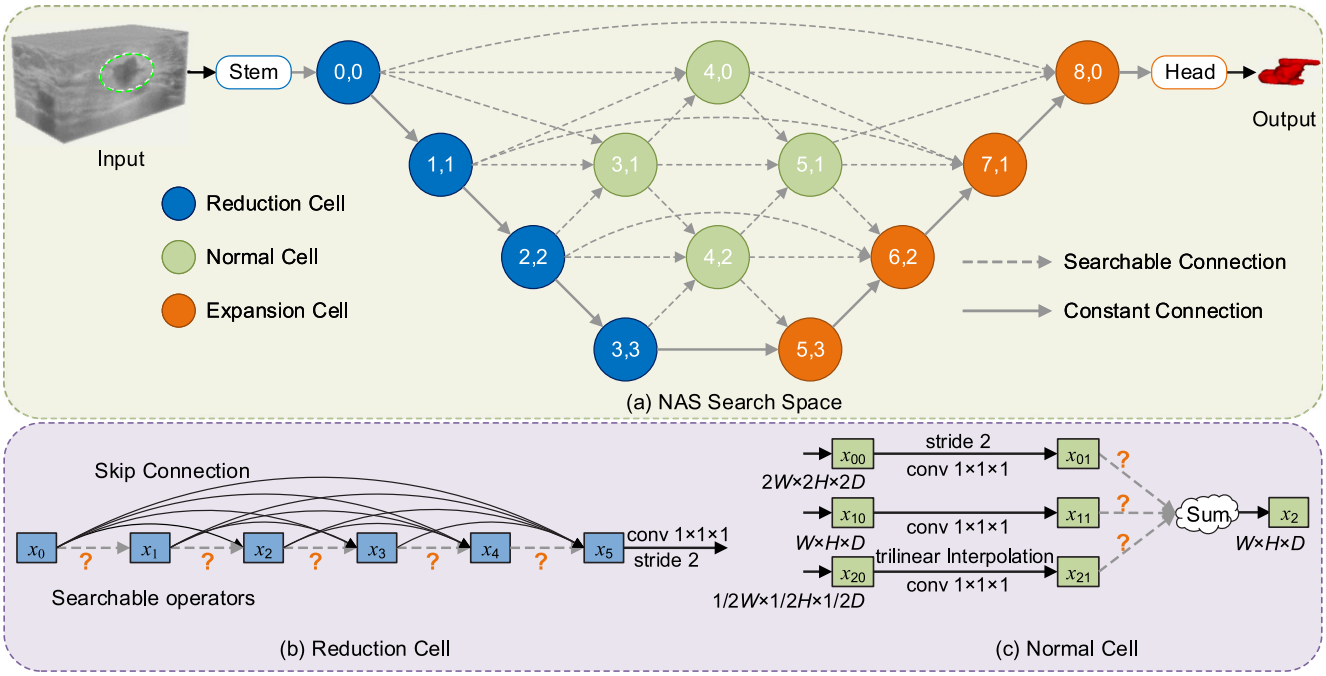


**Fig. 2.** Search space of the proposed Auto-DenseUNet for ABUS mass segmentation. (a) The overall search space, and (b–c) details of the reduction cell and the normal cell, respectively.

these data into 120 volumes (from 75 patients) for training and other 50 volumes (from 32 patients) for testing. More detailed information of the training and testing set is listed in Table 1, including the types of breast masses and the volume sizes of masses in the ABUS dataset. From (0, 500] $mm^3$ to (3000, 20 000] $mm^3$, the approximate maximum diameter of lesions in different groups are 1.2 cm, 1.8 cm, 2.6 cm and 4.8 cm, respectively. As the volume size variation is large, we divide the ABUS dataset into four groups according to the volume sizes of breast masses, as illustrated in Table 1. All volumes have an image size of $318 \times 716 \times 565$, with a voxel spacing of $0.52\ mm \times 0.07\ mm \times 0.21\ mm$ along transversal, coronal and sagittal views, respectively. As CNNs are not able to naively discover voxel spacing (Isensee et al., 2021), we first resample each ABUS volume with an isotropic voxel spacing of $0.52\ mm \times 0.52\ mm \times 0.52\ mm$, thus yielding a resampled image size of $318 \times 76 \times 290$. Then, we crop a $128 \times 64 \times 128$ region of interest (ROI) from each resampled volume before inputting it to the segmentation network. In most cases, the center of the ROI is the center of breast mass. If a breast mass is located near the border of the volume, the center of the ROI is adjusted far from the boundary. In this way, all the extracted ROIs have the same volume size with non-zero values, and the segmentation network will not memorize the locations of breast masses.

As there is no public 3D breast ultrasound dataset so far, we further utilize a publicly available 3D cardiac MRI dataset (left atrium segmentation) from the Medical Segmentation Decathlon challenge (Simpson et al., 2019) to evaluate the proposed Auto-DenseUNet. The 3D cardiac MRI dataset has 20 volumes for training and other 10 volumes for testing with large variability. Since the ground-truth labels for testing volumes are not provided in the 3D cardiac MRI dataset, we conducted a same five-fold cross-validation splits on the training dataset as the nnUNet. For a fair comparison, all the comparison methods utilize this same splits.

### 3.2. Methods

In this section, we first elaborate the proposed search space that defines all candidate network architectures. Then, we introduce a gradient-based search method, called continuous relaxation, which searches the optimal architecture in the search space. Finally, we present a novel decoupled search training strategy to relieve the memory limitation problem in the NAS-based medical image segmentation task.

#### 3.2.1. Search space

We follow the successful human-designed architecture and construct a searchable U-shape segmentation network with the encoder and decoder (Auto-DenseUNet). As shown in Fig. 2(a), the search space consists of encoding path (four reduction cells), decoding path (four expansion cells), and aggregation nodes (four normal cells), where a cell is defined as small repeated modules. To encourage the information and gradient flow in the search space, we employ the densely-connected structure in the reduction cell (Fig. 2(b)). We further use several multiscale connections between each aggregation node in the search space to encourage multiscale information fusion (See dash arrows in Fig. 2(a)).

**Reduction Cell:** Reduction cell is expected to learn rich feature representations in the current scale and reduce the feature map size for the next cell. To this end, we design the reduction cell as a densely-connected structure, as illustrated in Fig. 2(b). Similar to the dense block in DenseNet (Huang et al., 2017), the densely-connected structure of a reduction cell can be formulated as:

$$x_l = \mathcal{H}_l([x_0, x_1, \dots, x_{l-1}]) \tag{1}$$

where $x_l$ denotes the $l$th feature map in the reduction cell. $l \in L$ is the layer number, and $[\cdot]$ is the concatenation function. $\mathcal{H}$ is the non-linear transformation, which is defined as a combination of convolution (Conv), batch normalization (BN), and rectified linear units (ReLu). The 'conv' in $\mathcal{H}$ is searchable (denoted as gray dotted lines in Fig. 2(b)), with the following candidate operators: $3 \times 3 \times 3$ conv, $3 \times 3 \times 3$ conv with dilation 2, $3 \times 3 \times 3$ conv with dilation 3, $5 \times 5 \times 5$ conv, and $5 \times 5 \times 5$ conv with dilation 2.

According to Eq. (1), the number of feature maps in the $l$th layer is $k_0 + k \times (l-1)$, where $k_0$ is the number of input channels, and $k$ is called growth rate. In this paper, we adopt three approaches to improve the computational efficiency. First, we set a smaller layer $L$

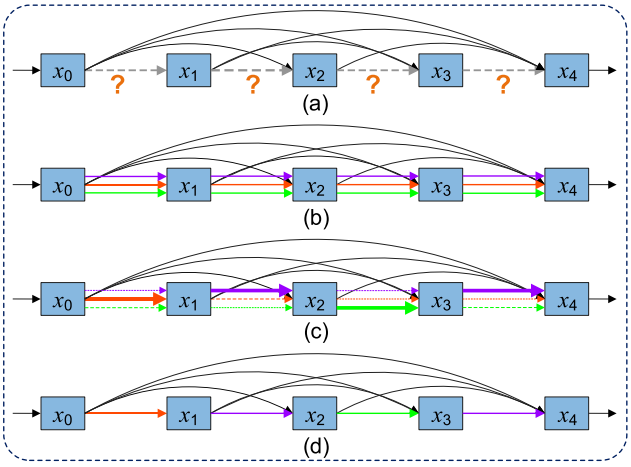


**Fig. 3.** Graphical illustration of the continuous relaxation approach. (a) A densely-connected cell structure to be learned. Dashed arrows represent searchable operations, and the specific operator is unknown. (b) Continuous relaxation of the search space. Each edge between two adjacent feature maps is a mixture of all candidate operators. (c) Joint optimization of network weights and architectures. Thicker arrows indicate larger weights for respective operators. (d) Discrete search space. The final architecture is determined by selecting operators with the largest weights.

in reduction cell. $L$ was set as 6 for all the reduction cells. Second, a bottleneck layer is employed into the reduction cell. Concretely, we utilize a $1 \times 1 \times 1$ convolution before the non-linear transformation $\mathcal{H}$ to produce 4k feature maps. In this way, the number of feature maps in deeper layers can be reduced. Finally, the output of reduction cell is further followed by a $1 \times 1 \times 1$ convolution, and the number of feature maps is reduced as half of the last layer in dense block. A stride of 2 is set in the convolution to downsample the feature map size by 2.

**Normal Cell:** The four normal cells (green circles in Fig. 2(a)) are designed for information fusion across different scales of feature maps, since multiscale feature fusion is proven to be effective in medical image segmentation tasks by human-designed networks (Zhou et al., 2020). As shown in Fig. 2(a), each normal cell takes feature maps from previous cells as input, and outputs to adjacent cells in different scales. Details inside the normal cell are illustrated in Fig. 2(c). A $1 \times 1 \times 1$ convolution is first utilized to convert feature maps from different scales into the same size and depth. Next, these feature maps are fed into searchable operators with different weights, and the output of normal cell is the weighted summation of three branches. The candidate operators in normal cell are $3 \times 3 \times 3$ conv, $3 \times 3 \times 3$ conv with dilation 2, $3 \times 3 \times 3$ conv with dilation 3, and $5 \times 5 \times 5$ conv. It is worth noting that the weights of three branches are also learnable during searching. When searching is completed, if the weight is below a certain threshold, the whole corresponding branch will be deleted from the network architecture.

**Expansion Cell:** The main function of expansion cell is to restore feature maps to the original image size. Besides, the expansion cell is not desired to learn abundant feature representations as the reduction cell and the normal cell. Therefore, all the operations in expansion cells are fixed. Concretely, we first construct the expansion cell similarly as the normal cell, and then replace the searchable operators with trilinear interpolation, followed by a fixed $3 \times 3 \times 3$ convolution. The weight of constant connection in expansion cell is set as 1 in order to assign larger weight on the decoder part of the network.

**Stem and Head:** In segmentation task, the stem is usually designed as a group of Conv-ReLu-BNs to reduce the size of feature maps, so that we can build a larger search space under the limited computational memory. Consequently, we follow the existing structures (Cao et al., 2021), and set the stem as a $7 \times 7 \times 7$ convolution with stride 2, followed by a $3 \times 3 \times 3$ max-pooling operation. The head includes two trilinear interpolations followed by convolutions to ensure that the output of segmentation network has the same dimension as the input.

**Algorithm 1** Pseudo code of the bi-level optimization process

1: **while** not converge **do**
2: Update network weights $w$ by $\nabla_w \mathcal{L}_{trainA}(w, \alpha, \beta)$
3: Update architecture $\alpha, \beta$ by $\nabla_{\alpha,\beta} \mathcal{L}_{trainB}(w^*(\alpha, \beta), \alpha, \beta)$
4: **end while**
5: Construct the discrete architecture according to $\alpha$ and $\beta$

### 3.2.2. Continuous relaxation of search space

In this paper, we employ the continuous relaxation approach (Liu et al., 2019b) to transform the discrete search space into continuous space, so that the architecture can be searched through a gradient-based optimization. Two types of architectures can be searched, including cell architecture and network architecture. The cell architecture is defined as searchable operations inside cells, while network architecture represents connections between different cells.

For the cell architecture, the continuous relaxation can be formulated as:

$$\bar{o}^{(i,j)}(x) = \sum_{o \in O} \frac{\exp\left(\alpha_o^{(i,j)}\right)}{\sum_{o' \in O} \exp\left(\alpha_{o'}^{(i,j)}\right)} o(x) \tag{2}$$

where $x$ represents the potential feature maps. $o \in O$ represents a non-linear operation on $x$, and $O$ denotes all the available operations in the current cell. $\alpha_o^{(i,j)}$ indicates a learnable weight on the current operation (from feature layer $i$ to feature layer $j$). In this way, the unknown operation on the current feature map is represented as a weighted combination of all available operations, and the task of finding the best operation in search space is transformed into learning of weights $\alpha$. In the end of search, the optimal operation can be found by selecting operator with the largest $\alpha$, as shown in Eq. (3). Fig. 3 shows a graphic illustration of the continuous relaxation.

$$o^{(i,j)} = \text{argmax}_{o \in O}\, \alpha_o^{(i,j)} \tag{3}$$

Similar to the cell architecture, we design the continuous relaxation of network architecture as:

$$\bar{x} = \sum_{c \in C} \frac{\exp\left(\beta_c\right)}{\sum_{c' \in C} \exp\left(\beta_{c'}\right)} T(x_c) \tag{4}$$

where $\bar{x}$ represents the output feature maps in the current cell. $x_c$ denotes the output feature maps in the previous cells, and $C$ indicates all the cells that connect to the current cell. $T$ denotes non-linear transformation, including feature maps aligned in different scales and the mixed operators $\bar{o}^{(i,j)}(x)$ in Eq. (2). In the end of searching, a searchable connection is reserved if its corresponding weight $\beta_c$ is larger than a certain threshold.

During the searching, the training set is divided into two parts: *train*A and *train*B. Then, the network architecture ($\alpha$, $\beta$) and network parameters ($w$) can be trained by solving the following bi-level optimization problem (Liu et al., 2019a):

$$\begin{aligned} \min_{\alpha,\beta} \quad & \mathcal{L}_{trainB}(w^*(\alpha, \beta), \alpha, \beta) \\ \text{s.t.} \quad & w^*(\alpha, \beta) = \text{argmin}_w \mathcal{L}_{trainA}(w, \alpha, \beta) \end{aligned} \tag{5}$$

As illustrated in Algorithm 1, the optimization process in Eq. (5) can be conducted by alternatively updating network weights $w$ and architectures $\alpha, \beta$. In the end of searching, a discrete segmentation network can be constructed according to the learned architectures $\alpha$ and $\beta$. In this paper, we employ Dice loss (Milletari et al., 2016) as loss function for breast mass segmentation to avoid biased predictions to background, and also employ a first-order approximation approach (Liu et al., 2019b) in Eq. (5) to improve the optimization efficiency.

#### 3.2.3. Decoupled search training

The continuous relaxation of search space consumes most of the computational memory during the searching, which limits the network with a larger number of feature maps. The reason is that the continuous relaxation combines all the feature maps of available operators. Besides, the 3D medical image segmentation task is another reason that contributes to larger memory usage. Generally, the NAS method first searches the optimal architecture in the search space, and then re-trains the searched architecture from scratch. Meanwhile, most existing methods will not change every detail of the searched network architecture. Since the NAS can be seen as a special case of hyper-parameter tuning (Liu et al., 2019b), if the number of feature maps is treated as a hyper-parameter (independent of NAS), we could search the architecture with a smaller number of feature maps, and train the searched network with a larger number of feature maps. Based on this observation, we introduce a decoupled search training strategy to alleviate the memory limitation problem.

Concretely, as the reduction cells are densely-connected structure, the number of feature maps of each cell can be calculated exactly. Then, we set normal cells and expansion cells to have the same number of feature maps as the reduction cell in the same level. For example, $\text{cell}^{3,1}$, $\text{cell}^{5,1}$ and $\text{cell}^{7,1}$ have the same feature map numbers as the output of $\text{cell}^{1,1}$. In this way, feature map numbers in the whole search space are only determined by the number of initial features of $\text{cell}^{0,0}$ ($N_{x0}$) and growth rate ($k$). In this paper, we set $N_{x0}$ and $k$ as two separate hyper-parameters, and use smaller $N_{x0}$ and $k$ during architecture search. Then, we train the optimal network with larger $N_{x0}$ and $k$. During the searching, $N_{x0}$ and $k$ are set to 64 and 16, respectively, according to the trade-off between memory usage and computational efficiency.

## 4. Experiments and results

### 4.1. Implementation details

All the experiments in this paper are conducted using PyTorch with an NVIDIA 2080Ti GPU. In the searching process, Adam (Kingma and Ba, 2015) optimizer is employed for the optimization of network weights $w$. Both initial learning rate and weight decay are set as 1e–3, and cosine learning rate scheduler is adopted. Network architectures $\alpha$ and $\beta$ are initialized as standard Gaussian times 1e–3, and optimized with another Adam optimizer. The initial learning rate is set as 1e–3, and weight decay is set as 1e–4. Following the work in Liu et al. (2019a), we start optimizing $\alpha$ and $\beta$ after 20 training epochs to avoid local optima of architectures when the network weights $w$ are not well trained. In both searching and training, the batch size is set as 2, and data augmentation strategy is utilized on-the-fly, including random flip and elastic transform. We search all the operators independently in reduction cells to encourage a more flexible search space. In addition, same as existing modular search strategies (Saikia et al., 2019; Li et al., 2021), all normal cells share the same operator to reduce the complexity of search space to some degree. In the end of searching process, we first build the discrete segmentation network according to the searched architecture, and then re-train the searched segmentation network from scratch. To better determine those hyper-parameters, we split 20 volumes (from 13 patients) in the training set for validation. For the cell architecture, we select the operators with the largest weights (Eq. (3)). For the network architecture, we threshold the connection weights with 0, 0.2, 0.4, 0.6, 0.8 and 1, and then select the architecture with the best performance. In this paper, the optimal architecture occurs when threshold is 0. Thus, we preserve all the connections between different cells. The searching time is around 30 h, while the training process takes around 10 h.

**Table 2**
Searched operators in reduction cells. $L$ is the layer number in these cells.

| $L$ | Cell | | | |
|---|---|---|---|---|
| | (0,0) | (1,1) | (2,2) | (3,3) |
| 1 | 5 × 5 × 5 conv with dilation 2 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv with dilation 2 |
| 2 | 5 × 5 × 5 conv with dilation 2 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv | 3 × 3 × 3 conv with dilation 3 |
| 3 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv with dilation 2 | 3 × 3 × 3 conv | 3 × 3 × 3 conv with dilation 2 |
| 4 | 3 × 3 × 3 conv | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv | 3 × 3 × 3 conv with dilation 3 |
| 5 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv with dilation 3 | 5 × 5 × 5 conv with dilation 2 | 3 × 3 × 3 conv with dilation 3 |
| 6 | 5 × 5 × 5 conv with dilation 2 | 3 × 3 × 3 conv with dilation 3 | 3 × 3 × 3 conv with dilation 3 | 5 × 5 × 5 conv |

**Table 3**
Testing results with different number of initial feature maps and different growth rate after neural architecture search. ($\text{Mean}_{\text{Std}}$).

| $k$ | $N_{x0}$ | DSC (%) | JI (%) | HD (mm) | Pre (%) | Rec (%) |
|---|---|---|---|---|---|---|
| 16 | 64 | $74.1_{16.3}$ | $61.0_{16.4}$ | $6.63_{5.81}$ | $79.2_{18.1}$ | $72.1_{18.3}$ |
| 32 | 64 | $75.4_{12.3}$ | $61.9_{13.4}$ | $6.45_{4.78}$ | $77.5_{15.6}$ | $76.7_{14.4}$ |
| 32 | 32 | $74.4_{14.4}$ | $61.1_{16.1}$ | $4.93_{2.85}$ | $\mathbf{81.2_{14.0}}$ | $72.7_{19.2}$ |
| 48 | 64 | $\mathbf{77.8_{10.3}}$ | $\mathbf{64.6_{12.4}}$ | $\mathbf{4.63_{2.41}}$ | $80.2_{14.5}$ | $\mathbf{78.2_{12.4}}$ |
| 64 | 64 | $76.4_{11.8}$ | $63.0_{13.3}$ | $5.07_{3.22}$ | $79.2_{14.6}$ | $75.8_{13.7}$ |

**Table 4**
Testing segmentation results with and without multi-scale aggregation. ($\text{Mean}_{\text{Std}}$).

| Multi-scale aggregation | DSC (%) | JI (%) | HD (mm) | Pre (%) | Rec (%) |
|---|---|---|---|---|---|
| Without | $74.5_{14.7}$ | $61.2_{15.9}$ | $4.93_{3.01}$ | $\mathbf{83.7_{11.7}}$ | $71.1_{19.1}$ |
| With | $\mathbf{77.8_{10.3}}$ | $\mathbf{64.6_{12.4}}$ | $\mathbf{4.63_{2.41}}$ | $80.2_{14.5}$ | $\mathbf{78.2_{12.4}}$ |

### 4.2. Experimental methods

First, ablation studies are carried out to verify the effectiveness of the proposed NAS method. Then, extensive comparison is conducted between the searched architecture and the existing networks on 3D medical image segmentation task. Finally, the proposed method is compared with existing segmentation methods in both the ABUS dataset and the cardiac MRI dataset.

Five evaluation metrics are employed to compare different segmentation methods, including the region-based Dice similarity coefficient (DSC) and Jaccard index (JI), boundary-based Hausdorff distance (HD), as well as the pixel-wise precision (Pre) and pixel-wise recall (Rec). The DSC and JI are defined as:

$$\text{DSC} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}, \ \text{JI} = \frac{TP}{TP + FP + FN} \tag{6}$$

where TP, FP, TN, FN are true positive, false positive, true negative and false negative, respectively. The Hausdorff distance is formulated as:

$$\text{HD}(P, G) = \max \left\{ \max_{S_P \in S(P)} d\left(S_P, S(G)\right), \max_{S_G \in S(G)} d\left(S_G, S(P)\right) \right\} \tag{7}$$

where $S(P)$ denotes the surface voxels of network predictions $P$, and $S(G)$ denotes the surface voxels of ground truth $G$. $d(S_G, S(P))$ represents the shortest distance between an arbitrary voxel in $S_G$ and $S(P)$, and vice versa. The pixel-wise precision and recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

### 4.3. Ablation studies

Table 2 lists searched operators in the four reduction cells. It can be observed that atrous convolution (convolution with dilation) is extensively used. The reason may be that atrous convolution is able to

**Table 5**
Comparison between searched operators and commonly-used operators by human. ($Mean_{Std}$).

| Kernel | DSC (%) | JI (%) | HD (mm) | Pre (%) | Rec (%) |
|---|---|---|---|---|---|
| 3 × 3 conv | $74.2_{17.5}$ | $61.3_{16.8}$ | $6.83_{5.90}$ | $78.8_{19.2}$ | $72.5_{19.2}$ |
| 3 × 3 conv, dilation 2 | $72.9_{12.5}$ | $58.7_{14.3}$ | $6.88_{5.16}$ | $79.0_{14.8}$ | $70.9_{16.4}$ |
| 5 × 5 conv | $73.1_{15.9}$ | $59.5_{15.9}$ | $6.77_{4.48}$ | $78.7_{17.2}$ | $71.3_{18.7}$ |
| Searched operators | $\mathbf{77.8}_{10.3}$ | $\mathbf{64.6}_{12.4}$ | $\mathbf{4.63}_{2.41}$ | $\mathbf{80.2}_{14.5}$ | $\mathbf{78.2}_{12.4}$ |

**Table 6**
The p-values from the student's t-test between searched operators and commonly-used operators used by human.

| Kernel | DSC | JI | HD | Pre | Rec |
|---|---|---|---|---|---|
| Searched vs. 3 × 3 conv | 4.17e−2 | 2.86e−2 | 5.89e−3 | 0.400 | 7.77e−3 |
| Searched vs. 3 × 3 conv, dilation 2 | 6.90e−5 | 2.67e−5 | 5.86e−3 | 0.320 | 8.86e−6 |
| Searched vs. 5 × 5 conv | 1.11e−2 | 2.29e−3 | 3.25e−3 | 0.394 | 5.47e−4 |

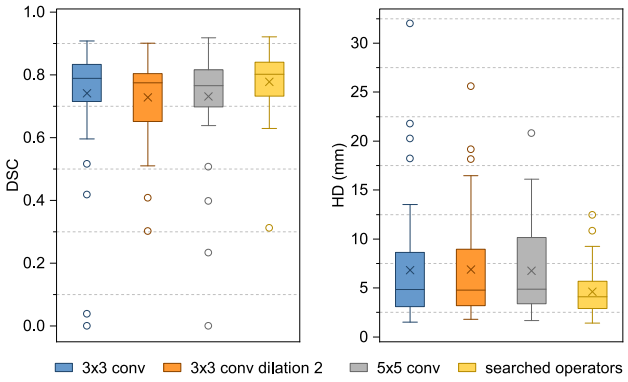


**Fig. 4.** Boxplot of different network architectures under different operators.

cover larger receptive field with the same network weights as normal convolution (Chen et al., 2018). Consequently, the network is encouraged to extract global structural representations of large masses. Since we adopt a modular search strategy, all normal cells share a searched 1 × 1 × 1 conv. This is consistent with our intuitive understanding, as the main function of normal cells is to improve information fusion from different feature maps, instead of learning more feature representations like reduction cells.

As mentioned in Section 3.2.3, the mixed operators in search space prevent the network from using a large number of feature maps due to the continuous relaxation. Therefore, we set the growth rate $k$ to 16 and the number of initial feature maps $N_{x0}$ to 64 during the searching, and then utilize larger $k$ and $N_{x0}$ to train the searched architecture. In this way, the segmentation network is encouraged to learn better representations with a larger number of feature maps and also optimal operators and connections. Table 3 lists the testing segmentation results under different number of initial feature maps as well as different growth rate. It can be seen that the segmentation performance increases when these two hyper-parameters increase, and then decreases slightly when both $k$ and $N_{x0}$ reach 64. The best segmentation results occur when $k$ and $N_{x0}$ are set to 48 and 64, respectively, with DSC of 77.8%, JI of 64.6%, and HD of 4.63 mm. This result is obviously better than the baseline method ($k = 16$, $N_{x0} = 64$), which indicates the effectiveness of the decoupled search training strategy. In the following experiments, all the architectures utilize $k = 48$, $N_{x0} = 64$ during training.

To validate the effectiveness of the multiscale feature aggregation scheme, we compare architectures with and without normal cells. Specifically, we build a network by removing all normal cells from the search space, and just preserve the skip connections between reduction cell and expansion cell. Table 4 lists quantitative testing results with and without the multiscale aggregation scheme. As shown in Table 4, when adding the multiscale information fusion, the segmentation performance improves with respect to DSC, JI and HD. Although the network without multiscale fusion achieves better precision (83.7%), its recall is just 71.1%. This means that more breast mass regions are missed than the network with multiscale information fusion.

To further test the performance of the searched architecture, we replaced all the searched operators with commonly-used convolutional kernels by human designer. These commonly-used operators include 3 × 3 × 3 conv, 5 × 5 × 5 conv, and 3 × 3 × 3 conv with dilation 2. As illustrated in Table 5, the searched architecture achieves a significant improvement in the testing set compared with other methods, in terms of all evaluation metrics. Table 6 lists p-values from student's t-test between the searched operators and the commonly-used operators by human designers. As shown in Table 6, except for the precision (Pre), all other evaluation metrics are statistically significant ($p < 0.05$).

Fig. 4 shows boxplot of DSC and HD under different network architectures with different operators. As can be seen in this figure, the searched architecture has the best median and mean values compared with the architectures with fixed operator. Particularly, the HD of the searched architecture is significantly lower than other architectures, indicating that the boundary of the prediction result is closer to the ground truth. Fig. 5 shows four cases of segmentation results using different architectures. The green, orange, blue, cyan and red lines are the ground truth, as well as segmentation results using 3 × 3 × 3 conv, 3 × 3 × 3 conv with dilation 2, 5 × 5 × 5 conv, and the searched operators, respectively. It can be observed that the searched architecture achieves the best agreement with the ground truth.

### 4.4. Comparison with different segmentation networks

We compare our Auto-DenseUNet with several classical segmentation networks on the 3D ABUS mass segmentation, including the 3D UNet (Çiçek et al., 2016), 3D ResUNet (Lee et al., 2017), VNet (Milletari et al., 2016), $D^2$UNet (Cao et al., 2021), and the nnUNet (Isensee et al., 2021). For a fair comparison, we train the nnUNet using the original implementation to ensure the optimal performance of nnUNet, and evaluate the testing set using a single model without model ensembling. All the other segmentation networks were derived from the available implementations and trained using the same configurations as our proposed Auto-DenseUNet, except for the network architectures. Table 7 compares segmentation results on testing data using our Auto-DenseUNet and other networks. It can be observed that the Auto-DenseUNet outperforms human-designed backbones considerably. Our method achieves an improvement of over 10% than 3D UNet and 3D ResUNet, and over 7% than VNet, in terms of DSC. Furthermore, an obvious decrease of standard deviation compared with existing backbones can also be seen. It is worth noting that the $D^2$UNet is carefully designed specially for the ABUS mass segmentation task, in order to address large shape and size variations of breast masses. However, our method still performs better than the $D^2$UNet, with an improvement of 4.2% and 3.5% in terms of DSC and JI, respectively. The nnUNet achieves better average DSC and JI than the proposed Auto-DenseUNet. However, the standard deviation of DSC and JI of nnUNet is higher than ours. In addition, the proposed Auto-DenseUNet achieves smaller HD than nnUNet, indicating that the Auto-DenseUNet has better surface similarity with the ground truth.

For Table 7, we design student's t-test between our method and other segmentation networks. As shown in Table 8, except for nnUNet, most of the evaluation metrics are statistically significant ($p < 0.05$) between the proposed method and existing methods. Except for Pre ($p < 0.05$), there is no statistical significance between the proposed method and the nnUNet, for most of the evaluation metrics.

Fig. 6 shows four cases of 3D segmentation results on the testing data using different backbones. From the first to the last columns are the original images, segmentation results of 3D UNet, 3D ResUNet,

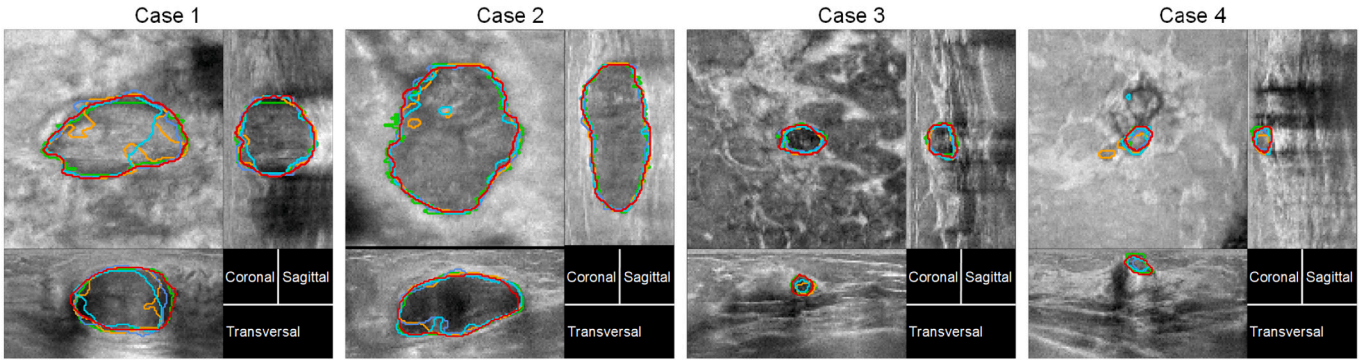


**Fig. 5.** Segmentation results on testing data using different architectures. Green, orange, blue, cyan and red lines are the ground truth, as well as the segmentation results using $3 \times 3 \times 3$ conv, $3 \times 3 \times 3$ conv with dilation 2, $5 \times 5 \times 5$ conv, and searched operators, respectively.

**Table 7**
Comparison with other segmentation networks. ($Mean_{Std}$).

| Network | Params | Training time | DSC (%) | JI (%) | HD (mm) | Pre (%) | Rec (%) |
|---|---|---|---|---|---|---|---|
| 3D UNet (Çiçek et al., 2016) | 19.1M | 5.4 h | $64.5_{25.3}$ | $52.0_{23.7}$ | $11.0_{7.44}$ | $76.9_{25.7}$ | $61.7_{29.0}$ |
| 3D ResUNet (Lee et al., 2017) | 19.5M | 5.8 h | $65.4_{24.7}$ | $52.8_{23.2}$ | $10.1_{7.57}$ | $76.3_{26.6}$ | $61.4_{27.2}$ |
| VNet (Milletari et al., 2016) | 45.6M | 7.7 h | $70.6_{20.4}$ | $57.7_{20.0}$ | $7.69_{5.81}$ | $79.5_{18.8}$ | $69.7_{24.9}$ |
| $D^2$UNet (Cao et al., 2021) | 2.0M | 11.3 h | $73.6_{19.4}$ | $61.1_{19.6}$ | $5.94_{5.28}$ | $80.3_{21.3}$ | $72.8_{21.2}$ |
| nnUNet (Isensee et al., 2021) | 30.8M | 45.5 h | $\mathbf{78.1_{14.1}}$ | $\mathbf{65.9_{16.1}}$ | $5.54_{4.05}$ | $\mathbf{86.1_{13.7}}$ | $75.4_{17.4}$ |
| Ours | 20.0M | 9.9 h | $77.8_{10.3}$ | $64.6_{12.4}$ | $\mathbf{4.63_{2.41}}$ | $80.2_{14.5}$ | $\mathbf{78.2_{12.4}}$ |

**Table 8**
The p-values from the student's t-test between our methods and other segmentation networks.

| Network | DSC | JI | HD | Pre | Rec |
|---|---|---|---|---|---|
| Ours vs. 3D UNet (Çiçek et al., 2016) | 4.32e−5 | 1.77e−5 | 4.28e−7 | 0.287 | 8.33e−6 |
| Ours vs. 3D ResUNet (Lee et al., 2017) | 2.04e−4 | 1.05e−4 | 1.15e−5 | 0.271 | 4.74e−6 |
| Ours vs. VNet (Milletari et al., 2016) | 5.79e−3 | 4.10e−3 | 1.77e−4 | 0.784 | 2.84e−3 |
| Ours vs. $D^2$UNet (Cao et al., 2021) | 3.48e−2 | 0.502 | 3.98e−2 | 0.397 | 6.96e−3 |
| Ours vs. nnUNet (Isensee et al., 2021) | 0.836 | 0.491 | 0.207 | 4.74e−7 | 0.175 |

VNet, $D^2$UNet, nnUNet and Auto-DenseUNet, and the ground truth. Green circle in the first column shows approximate location of breast mass. As shown in Fig. 6, in most cases, our method achieves better segmentation results compared with other backbones, with less errors in both large and small breast masses.

To test the generalization of our proposed method and other methods on different mass sizes, we divide the testing data into four groups according to the mass volume size, as shown in Table 1. Fig. 7 shows boxplots of HD under different volume size groups and different backbones in the testing data set. It can be seen that our method performs better than other methods in both groups of (0, 500] $mm^3$ and (1000, 3000] $mm^3$, with greater mean value and smaller standard deviation. In addition, in most cases, our method has less outliers, which are often caused by false positive regions in the segmentation results.

### 4.5. Comparison with existing methods in the ABUS dataset

In this section, we also compare our Auto-DenseUNet with existing methods on ABUS breast mass segmentation task, as shown in Table 9. 95HD is 95% Hausdorff distance (Zhou et al., 2021a), which is often used to remove the influence of outliers in segmentation results. As there is no publicly available ABUS dataset, the comparison results in Table 9 are obtained from their own dataset. That is, these results are reported in their original article. As shown in Table 9, our proposed method achieves a competitive result compared with other methods on ABUS mass segmentation. Though the DSC is the same as (Zhou et al., 2021a), our method achieves a better 95HD (2.21 mm) than (Zhou et al., 2021a), indicating better boundary similarity with the ground truth. However, it is worth noting that the comparison in Table 9 is not very fair, since the dataset and the configurations used in these methods are different.

**Table 9**
Comparison with existing algorithms on ABUS mass segmentation.

| Method | Year | DSC | 95HD (mm) |
|---|---|---|---|
| Tan et al. (2016) | 2016 | 0.73 | – |
| Agarwal et al. (2017) | 2017 | 0.69 | – |
| Kozegar et al. (2018) | 2018 | 0.74 | – |
| Wang et al. (2020) | 2020 | 0.58 | – |
| Cao et al. (2021) | 2021 | 0.76 | 2.56 |
| Zhou et al. (2021a) | 2021 | 0.78 | 3.30 |
| Ours | – | 0.78 | 2.21 |

**Table 10**
Five-fold cross-validation results using different methods in the publicly available cardiac MRI dataset ($Mean_{Std}$). The notion '–' in Std means data are not provided in the reference paper.

| Methods | | DSC (%) |
|---|---|---|
| Human-designed | ResUNet (Lee et al., 2017) | 89.60 – |
| | VNet (Milletari et al., 2016) | $91.88_{1.91}$ |
| | $D^2$UNet (Cao et al., 2021) | $92.66_{1.94}$ |
| | nnUNet-fullres (Isensee et al., 2021) | 93.28 – |
| | nnUNet-ensemble (Isensee et al., 2021) | 92.68 – |
| NAS | SCNAS (Kim et al., 2019) | 91.29 – |
| | SCNAS (transfer) (Kim et al., 2019) | 91.91 – |
| | RONASMIS (Bae et al., 2019) | 92.72 – |
| | Auto-DenseUNet (Ours) | $92.88_{1.58}$ |

### 4.6. Comparison with existing methods in the cardiac dataset

To further verify the effectiveness of the proposed method, segmentation results of the proposed Auto-DenseUNet and the existing

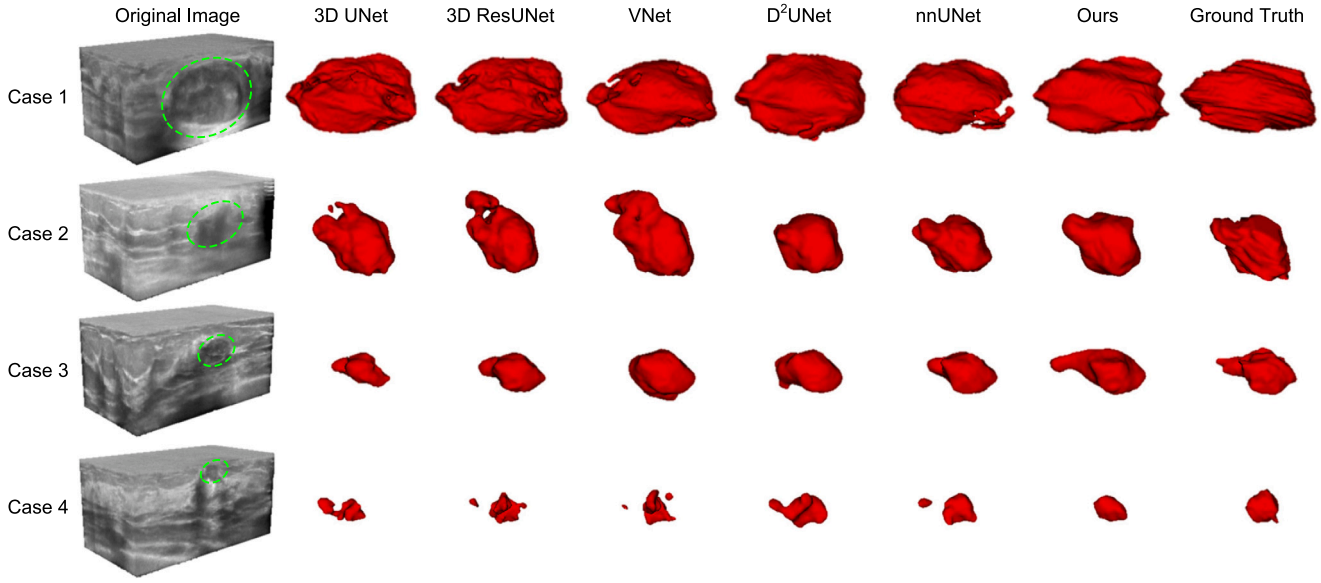


**Fig. 6.** Segmentation results using different networks.

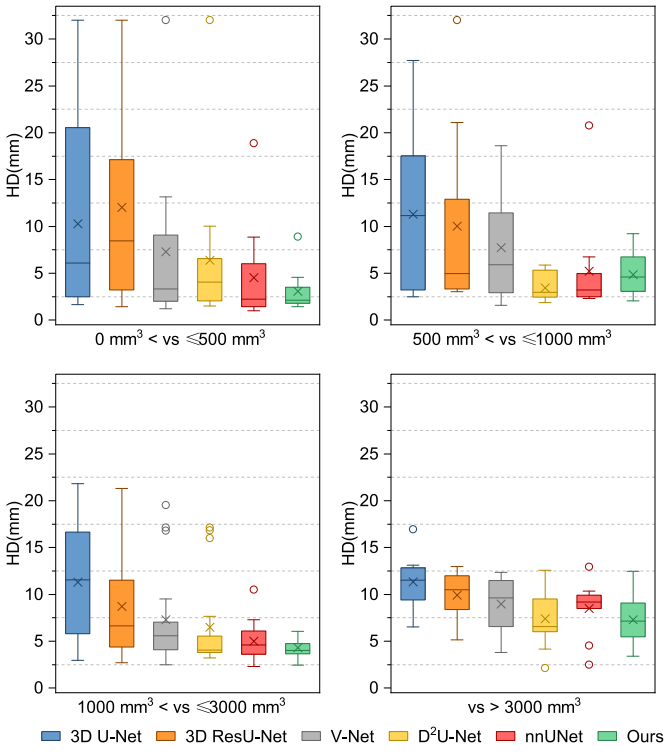


**Fig. 7.** Boxplot of HD under different volume size groups and different networks.

human-designed networks and NAS approaches on the publicly available cardiac MRI dataset are compared, shown in Table 10. Because labels of the testing data set in the cardiac MR dataset are not publicly available, we follow the existing works (Isensee et al., 2021; Kim et al., 2019; Bae et al., 2019) that evaluated segmentation performance with 5-fold cross validation on the training data set, not on the testing data set. For a fair comparison, results of nnUNet are reported in the original article, where nnUNet-fullres indicates nnUNet with 3D full resolutions, and nnUNet-ensemble indicates the best ensemble using the 2D U-Net and the 3D full resolution U-Net combination. All methods in Table 10 were trained with the same five-fold training-validation splits as nnUNet. We implement the VNet and DenseUNet, and train the two networks using the same training configurations as our Auto-DenseUNet. Segmentation results of the remaining human-designed networks are reported in the RONASMIS. The results of the NAS approaches are reported in their own article. It is worth noting that except for the nnUNet-ensemble method, no ensembling, testing time augmentation or post-processing are used to achieve final results. From Table 10, it can be seen that except for nnUNet-fullres, our proposed Auto-DenseUNet achieves the best DSC compared with the existing NAS and the human-designed networks, indicating the effectiveness of the proposed method.

## 5. Discussion

In recent years, deep learning based methods have made outstanding progress in medical image segmentation, and outperform traditional the hand-crafted-feature based algorithms significantly. However, designing segmentation network for a particular dataset is time-costing, and heavily relies on prior knowledge and experience of researchers. Consequently, a promising research direction to solve this problem is NAS, which aims to search the optimal architecture with the least intervention of human designers. Base on this observation, we study the NAS-based ABUS mass segmentation problem. As shown in Table 5, although the Auto-DenseUNet does not have regular operators like the human-designed network, the searched architecture outperforms other networks significantly.

We employ the densely-connected structure intensively in the search space, including the dense block in reduction cell (shown in Fig. 2(b)) and connections between two cells (shown in Fig. 2(a)). This is based on the success of many human-designed networks (Li et al., 2018; Dolz et al., 2019) and searched architecture by NAS (Ji et al., 2020). Besides, the motivation behind the densely-connected structure is due to several observations (Huang et al., 2017). First, direct connections between all feature maps improve information and gradient flows throughout the entire network. Second, the dense connectivity pattern has better parameter efficiency, since there is no need to re-learn redundant feature maps. Third, the dense connections have a regularization effect, which reduces the risk of over-fitting on tasks with smaller datasets. In the searched architecture, it can be seen that most connections between cells are preserved, indicating that the dense connections are necessary in our ABUS mass segmentation task.

Besides dense connections, information fusion from multi-levels is another factor that improves the performance of our method. Different from existing methods (Zhou et al., 2020; Ronneberger et al., 2015) that only add skip connection between same-scale feature maps, we explore feature fusion in different scales. Intuitively, the network is able to extract richer feature representations when aggregating information from different scales. As shown in Table 4, the Auto-DenseUNet achieves significant improvement when including the multiscale information fusion scheme, indicating effectiveness of our proposed method.

The contiguous relaxation makes it difficult to search on a large search space with high degree of freedom, as it consumes too much computational memory. This challenge is more severe in the ABUS mass segmentation due to the heavier computational burden brought by 3D volumes. With the proposed decoupled search training strategy, our method can put more emphasis on first searching the optimal cell architecture and network architecture. Then, the optimal feature map numbers can be determined in the searched architecture, as shown in Table 3. Experimental results in ablation studies prove the effectiveness of the decoupled search training method. This strategy makes our search space larger and deeper than the existing methods on 3D medical image segmentation. An alternative solution to this problem is to use the patch-based training strategy, where each volume is cropped into several patches. However, the drawback of patch-based approach is that it loses global structural information of ABUS volumes. Without global information, more false positives could be introduced to the segmentation result, since many normal tissues and artifacts have similar shapes and intensities as breast masses.

Although our proposed method does not outperform the nnUNet in most of metrics according to Table 7, the results in Table 8 shows no significant difference between our proposed method and the nnUNet. This indicate the results by our proposed method are similar to those by the nnUNet. In addition, from the boxplot in Fig. 7, it can be seen that the results by our proposed method are more stable in both breast mass groups of (0, 500] $mm^3$ and (1000, 3000] $mm^3$, with greater mean value and smaller standard deviation than those by the nnUNet. Nevertheless, many excellent rule-based configurations are not included in our search space, which are worth exploring in the NAS-based method. In the future, we will also learn experience from nnUNet to design a new search space with higher degree of freedom.

One possible limitation of our proposed Auto-DenseUNet is assuming same data distribution of the dataset used in the searching and training procedures. Actually, image distributions under different hospitals, different acquisition equipment, and different acquisition parameters could be quite different. Hence, the performance of the optimal architecture searched in one dataset may decrease when used in another dataset, even if these two datasets are of the same imaging modality. How to address the domain shift across different datasets in NAS is an interesting future direction to explore. Besides, we only segment breast masses in ROI, which may increase segmentation complexity in practice. This is because segmenting breast masses in the whole volume can face more challenge, such as random breast mass locations and limited GPU memory. In the future, we will design more effective networks to segment breast masses within the whole volume.

## 6. Conclusion

In this paper, we have designed an Auto-DenseUNet for breast mass segmentation in 3D ABUS images. Specifically, a novel search space with multi-scale feature aggregation scheme is constructed to encourage information fusion throughout the segmentation network. Also, a decoupled search training strategy is suggested to alleviate the computational memory limitation caused by the continuous relaxation searching approach and the 3D medical images. Quantitative and qualitative results on 50 testing ABUS volumes reveal that our proposed Auto-DenseUNet performs better than most of human-designed architectures, as well as existing methods, on ABUS mass segmentation. In the future, a search space with higher degree of freedom will be designed to further improve our proposed method, based on the successful experience of nnUNet.

## CRediT authorship contribution statement

**Xuyang Cao:** Conceptualization, Methodology, Software, Writing – original draft. **Houjin Chen:** Writing – review, Supervision. **Yanfeng Li:** Methodology, Writing – review & editing. **Yahui Peng:** Writing – review & editing, Supervision. **Yue Zhou:** Software, Formal analysis. **Lin Cheng:** Data curation. **Tianming Liu:** Conceptualization, Writing – review & editing. **Dinggang Shen:** Conceptualization, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## References


Agarwal, R., Diaz, O., Lladó, X., Gubern-Mérida, A., Vilanova, J.C., Martí, R., 2017. Lesion segmentation in automated 3D breast ultrasound: Volumetric analysis. Ultrason. Imaging 40 (2), 97–112. http://dx.doi.org/10.1177/0161734617737733.

Araújo, T., Abayazid, M., Rutten, M.J.C.M., Misra, S., 2017. Segmentation and three-dimensional reconstruction of lesions using the automated breast volume scanner (ABVS): Segmentation and 3D reconstruction of lesions using ABVS. Int. J. Med. Robot. Comput. Assist. Surg. 13 (3), 375–392. http://dx.doi.org/10.1002/rcs.1767.

Bae, W., Lee, S., Lee, Y., Park, B., Chung, M., Jung, K.-H., 2019. Resource optimized neural architecture search for 3D medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 228–236. http://dx.doi.org/10.1007/978-3-030-32245-8_26.

Barzegar, Z., Jamzad, M., 2019. Brain tumor segmentation based on 3D neighborhood features using rule-based learning. In: Eleventh International Conference on Machine Vision (ICMV 2018). 11041, International Society for Optics and Photonics, 1104103. http://dx.doi.org/10.1117/12.2523220.

Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., Cheng, L., 2021. Dilated densely connected U-net with uncertainty focus loss for 3D ABUS mass segmentation. Comput. Methods Programs Biomed. 106313. http://dx.doi.org/10.1016/j.cmpb.2021.106313.

Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., Cheng, L., 2021. Uncertainty aware temporal-ensembling model for semi-supervised ABUS mass segmentation. IEEE Trans. Med. Imaging 40 (1), 431–443. http://dx.doi.org/10.1109/TMI.2020.3029161.

Cao, X., Chen, H., Li, Y., Peng, Y., Zhou, Y., Cheng, L., 2020. Boundary loss with non-euclidean distance constraint for ABUS mass segmentation. In: 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). pp. 645–650. http://dx.doi.org/10.1109/CISP-BMEI51763.2020.9263676.

Chen, W., Gong, X., Liu, X., Zhang, Q., Li, Y., Wang, Z., 2020. Fasterseg: Searching for faster real-time semantic segmentation. In: 8th International Conference on Learning Representations, ICLR 2020. OpenReview.net, Addis Ababa, Ethiopia, pp. 1–14.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848. http://dx.doi.org/10.1109/TPAMI.2017.2699184.

Chu, J., Chen, Y., Zhou, W., Shi, H., Cao, Y., Tu, D., Jin, R., Xu, Y., 2020. Pay more attention to discontinuity for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 166–175. http://dx.doi.org/10.1007/978-3-030-59719-1_17.

Çiçek, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, Athens, Greece, pp. 424–432.

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., Ayed, I.B., 2019. HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation. IEEE Trans. Med. Imaging 38 (5), 1116–1126. http://dx.doi.org/10.1109/TMI.2018.2878669.

Gómez-Flores, W., Ruiz-Ortega, B.A., 2016. New fully automated method for segmentation of breast lesions on ultrasound based on texture analysis. Ultrasound Med. Biol. 42 (7), 1637–1650. http://dx.doi.org/10.1016/j.ultrasmedbio.2016.02.016.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. arXiv:1503.02531 URL: http://arxiv.org/abs/1503.02531.

Hospedales, T.M., Antoniou, A., Micaelli, P., Storkey, A.J., 2021. Meta-learning in neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 1. http://dx.doi.org/10.1109/TPAMI.2021.3079209.

Huang, G., Liu, Z., Maaten, L.v.d., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, Hawaii., pp. 2261–2269. http://dx.doi.org/10.1109/CVPR.2017.243.

Hutter, F., Kotthoff, L., Vanschoren, J. (Eds.), 2019. Automated Machine Learning: Methods, Systems, Challenges. In: The Springer Series on Challenges in Machine Learning, Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-05318-5.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnu-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211. http://dx.doi.org/10.1038/s41592-020-01008-z.

Ji, Y., Zhang, R., Li, Z., Ren, J., Zhang, S., Luo, P., 2020. UXNet: Searching multi-level feature aggregation for 3D medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 346–356. http://dx.doi.org/10.1007/978-3-030-59710-8_34.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78. http://dx.doi.org/10.1016/j.media.2016.10.004.

Kim, S., Kim, I., Lim, S., Baek, W., Kim, C., Cho, H., Yoon, B., Kim, T., 2019. Scalable neural architecture search for 3D medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 220–228. http://dx.doi.org/10.1007/978-3-030-32248-9_25.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015. San Diego, CA, USA, pp. 1–15, URL: http://arxiv.org/abs/1412.6980.

Kozegar, E., Soryani, M., Behnam, H., Salamati, M., Tan, T., 2018. Mass segmentation in automated 3-D breast ultrasound using adaptive region growing and supervised edge-based deformable model. IEEE Trans. Med. Imaging 37 (4), 918–928. http://dx.doi.org/10.1109/TMI.2017.2787685.

Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.-A., Li, J., Hu, Z., Wang, Y., Koohbanani, N.A., Jahanifar, M., Tajeddin, N.Z., Gooya, A., Rajpoot, N., Ren, X., Zhou, S., Wang, Q., Shen, D., Yang, C.-K., Weng, C.-H., Yu, W.-H., Yeh, C.-Y., Yang, S., Xu, S., Yeung, P.H., Sun, P., Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Smedby, O., Wang, C., Chidester, B., Ton, T.-V., Tran, M.-T., Ma, J., Do, M.N., Graham, S., Vu, Q.D., Kwak, J.T., Gunda, A., Chunduri, R., Hu, C., Zhou, X., Lotfi, D., Safdari, R., Kascenas, A., O'Neil, A., Eschweiler, D., Stegmaier, J., Cui, Y., Yin, B., Chen, K., Tian, X., Gruening, P., Barth, E., Arbel, E., Remer, I., Ben-Dor, A., Sirazitdinova, E., Kohl, M., Braunewell, S., Li, Y., Xie, X., Shen, L., Ma, J., Baksi, K.D., Khan, M.A., Choo, J., Colomer, A., Naranjo, V., Pei, L., Iftekharuddin, K.M., Roy, K., Bhattacharjee, D., Pedraza, A., Bueno, M.G., Devanathan, S., Radhakrishnan, S., Koduganty, P., Wu, Z., Cai, G., Liu, X., Wang, Y., Sethi, A., 2020. A multi-organ nucleus segmentation challenge. IEEE Trans. Med. Imaging 39 (5), 1380–1391. http://dx.doi.org/10.1109/TMI.2019.2947628.

Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S., 2017. Superhuman accuracy on the SNEMI3D connectomics challenge. arXiv:1706.00120 [Cs]. URL: http://arxiv.org/abs/1706.00120.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C., Heng, P., 2018. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans. Med. Imaging 37 (12), 2663–2674. http://dx.doi.org/10.1109/TMI.2018.2845918.

Li, Q., Wu, X., Liu, T., 2021. Differentiable neural architecture search for optimal spatial/temporal brain function network decomposition. Med. Image Anal. 69, 101974. http://dx.doi.org/10.1016/j.media.2021.101974.

Li, C., Xu, C., Gui, C., Fox, M.D., 2010. Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. 19 (12), 3243–3254. http://dx.doi.org/10.1109/TIP.2010.2069690.

Li, Q., Zhang, W., Lv, J., Wu, X., Liu, T., 2020. Neural architecture search for optimization of spatial-temporal brain network decomposition. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 377–386. http://dx.doi.org/10.1007/978-3-030-59728-3_37.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. http://dx.doi.org/10.1016/j.media.2017.07.005.

Liu, C., Chen, L.-C., Schroff, F., Adam, H., Hua, W., Yuille, A.L., Fei-Fei, L., 2019a. Auto-DeepLab: Hierarchical neural architecture search for semantic image segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 82–92. http://dx.doi.org/10.1109/CVPR.2019.00017.

Liu, L., Hu, X., Zhu, L., Fu, C.-W., Qin, J., Heng, P.-A., 2020a. $\phi$-Net: Stacking densely convolutional LSTMs for sub-cortical brain structure segmentation. IEEE Trans. Med. Imaging 39 (9), 2806–2817. http://dx.doi.org/10.1109/TMI.2020.2975642.

Liu, H., Simonyan, K., Yang, Y., 2019b. DARTS: Differentiable architecture search. arXiv:1806.09055. URL: http://arxiv.org/abs/1806.09055.

Liu, Z., Wang, H., Zhang, S., Wang, G., Qi, J., 2020b. NAS-SCAM: Neural architecture search-based spatial and channel joint attention module for nuclei semantic segmentation and classification. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 263–272. http://dx.doi.org/10.1007/978-3-030-59710-8_26.

Liu, C., Zoph, B., Neumann, M., Shlens, J., Hua, W., Li, L.-J., Fei-Fei, L., Yuille, A., Huang, J., Murphy, K., 2018. Progressive neural architecture search. In: Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 19–35. http://dx.doi.org/10.1007/978-3-030-01246-5_2.

Luo, G., 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics 5 (1), 18. http://dx.doi.org/10.1007/s13721-016-0125-6.

Mackay, M., Vicol, P., Lorraine, J., Duvenaud, D., Grosse, R., 2019. Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions. In: 7th International Conference on Learning Representations, ICLR 2019. OpenReview.net, New Orleans, LA, USA, pp. 1–25.

Milletari, F., Navab, N., Ahmadi, S., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision. Stanford, CA, USA, pp. 565–571. http://dx.doi.org/10.1109/3DV.2016.79.

Nekrasov, V., Chen, H., Shen, C., Reid, I.D., 2019. Fast neural architecture search of compact semantic segmentation models via auxiliary cells. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019. Computer Vision Foundation / IEEE, Long Beach, CA, USA, pp. 9126–9135. http://dx.doi.org/10.1109/CVPR.2019.00934.

Nie, D., Wang, L., Adeli, E., Lao, C., Lin, W., Shen, D., 2019. 3-d fully convolutional networks for multimodal isointense infant brain image segmentation. IEEE Trans. Cybern. 49 (3), 1123–1136. http://dx.doi.org/10.1109/TCYB.2018.2797905.

Real, E., Aggarwal, A., Huang, Y., Le, Q.V., 2019. Regularized evolution for image classifier architecture search. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. pp. 4780–4789. http://dx.doi.org/10.1609/aaai.v33i01.33014780.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., Wang, X., 2020. A comprehensive survey of neural architecture search: Challenges and solutions. arXiv:2006.02903. URL: http://arxiv.org/abs/2006.02903.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.

Ruan, Y., Li, D., Marshall, H., Miao, T., Cossetto, T., Chan, I., Daher, O., Accorsi, F., Goela, A., Li, S., 2020. MB-FSGAN: Joint segmentation and quantification of kidney tumor on CT by the multi-branch feature sharing generative adversarial network. Med. Image Anal. 64, 101721. http://dx.doi.org/10.1016/j.media.2020.101721.

Saikia, T., Marrakchi, Y., Zela, A., Hutter, F., Brox, T., 2019. AutoDispNet: Improving disparity estimation with autoML. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019. IEEE, Seoul, Korea (South), pp. 1812–1823. http://dx.doi.org/10.1109/ICCV.2019.00190.

Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J., 2019. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. pp. 1–15, arXiv:1902.09063 [Cs, Eess]. URL: http://arxiv.org/abs/1902.09063.

Tan, T., Gubern-Mérida, A., Borelli, C., Manniesing, R., van Zelst, J., Wang, L., Zhang, W., Platel, B., Mann, R.M., Karssemeijer, N., 2016. Segmentation of malignant lesions in 3D breast ultrasound using a depth-dependent model: Segmentation of malignant lesions in ABUS. Med. Phys. 43 (7), 4074–4084. http://dx.doi.org/10.1118/1.4953206.

Tan, T., Platel, B., Huisman, H., Sanchez, C.I., Mus, R., Karssemeijer, N., 2012. Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation. IEEE Trans. Med. Imaging 31 (5), 1034–1042. http://dx.doi.org/10.1109/TMI.2012.2184549.

Tian, Z., He, T., Shen, C., Yan, Y., 2019. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3121–3130. http://dx.doi.org/10.1109/CVPR.2019.00324.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2015. Learning spatiotemporal features with 3D convolutional networks. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). In: ICCV '15, IEEE Computer Society, Washington, DC, USA, pp. 4489–4497. http://dx.doi.org/10.1109/ICCV.2015.510.

van Zelst, J.C.M., Mann, R.M., 2018. Automated three-dimensional breast US for screening: Technique, artifacts, and lesion characterization. RadioGraphics 38 (3), 663–683. http://dx.doi.org/10.1148/rg.2018170162.

Vourtsis, A., Kachulis, A., 2018. The performance of 3D ABUS versus HHUS in the visualisation and BI-RADS characterisation of breast lesions in a large cohort of 1,886 women. Eur. Radiol. 28 (2), 592–601. http://dx.doi.org/10.1007/s00330-017-5011-9.

Wang, L., Nie, D., Li, G., Puybareau, E., Dolz, J., Zhang, Q., Wang, F., Xia, J., Wu, Z., Chen, J.-W., Thung, K.-H., Bui, T.D., Shin, J., Zeng, G., Zheng, G., Fonov, V.S., Doyle, A., Xu, Y., Moeskops, P., Pluim, J.P.W., Desrosiers, C., Ayed, I.B., Sanroma, G., Benkarim, O.M., Casamitjana, A., Vilaplana, V., Lin, W., Li, G., Shen, D., 2019a. Benchmark on automatic six-month-old infant brain segmentation algorithms: The iSeg-2017 challenge. IEEE Trans. Med. Imaging 38 (9), 2219–2230. http://dx.doi.org/10.1109/TMI.2019.2901712.

Wang, Y., Wang, N., Xu, M., Yu, J., Qin, C., Luo, X., Yang, X., Wang, T., Li, A., Ni, D., 2020. Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. IEEE Trans. Med. Imaging 39 (4), 866–876. http://dx.doi.org/10.1109/TMI.2019.2936500.

Wang, S., Yu, L., Li, K., Yang, X., Fu, C.-W., Heng, P.-A., 2019b. Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 102–110. http://dx.doi.org/10.1007/978-3-030-32239-7_12.

Xi, X., et al., 2017. Breast tumor segmentation with prior knowledge learning. Neurocomputing 237 (C), 145–157. http://dx.doi.org/10.1016/j.neucom.2016.09.067.

Xing, J., Li, Z., Wang, B., Qi, Y., Yu, B., Zanjani, F.G., Zheng, A., Duits, R., Tan, T., 2020. Lesion segmentation in ultrasound using semi-pixel-wise cycle generative adversarial nets. IEEE/ACM Trans. Comput. Biol. Bioinform. 1. http://dx.doi.org/10.1109/TCBB.2020.2978470.

Zhao, B., Chen, X., Li, Z., Yu, Z., Yao, S., Yan, L., Wang, Y., Liu, Z., Liang, C., Han, C., 2020. Triple U-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. Med. Image Anal. 65, 101786. http://dx.doi.org/10.1016/j.media.2020.101786.

Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.-T., Shen, D., 2021a. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med. Image Anal. 70, 101918. http://dx.doi.org/10.1016/j.media.2020.101918.

Zhou, Y., Chen, H., Li, Y., Wang, S., Cheng, L., Li, J., 2021b. 3D multi-view tumor detection in automated whole breast ultrasound using deep convolutional neural network. Expert Syst. Appl. 168, 114410. http://dx.doi.org/10.1016/j.eswa.2020.114410.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2020. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39 (6), 1856–1867. http://dx.doi.org/10.1109/TMI.2019.2959609.

Zoph, B., Le, Q.V., 2017. Neural architecture search with reinforcement learning. In: 5th International Conference on Learning Representations, ICLR 2017. OpenReview.net, Toulon, France, pp. 1–16, URL: https://openreview.net/forum?id=r1Ue8Hcxg.

Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2018. Learning transferable architectures for scalable image recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT, pp. 8697–8710. http://dx.doi.org/10.1109/CVPR.2018.00907.