



Research
AI for Precision Medicine—Review

Deep Learning in Medical Ultrasound Analysis: A Review

Shengfeng Liu^a, Yi Wang^a, Xin Yang^b, Baiying Lei^a, Li Liu^a, Shawn Xiang Li^a, Dong Ni^{a,*},
Tianfu Wang^{a,*}



^a National-Regional Key Technology Engineering Laboratory for Medical Ultrasound & Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging & School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China

^b Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:

Received 1 March 2018

Revised 18 July 2018

Accepted 8 November 2018

Available online 29 January 2019

Keywords:

Deep learning

Medical ultrasound analysis

Classification

Segmentation

Detection

ABSTRACT

Ultrasound (US) has become one of the most commonly performed imaging modalities in clinical practice. It is a rapidly evolving technology with certain advantages and with unique challenges that include low imaging quality and high variability. From the perspective of image analysis, it is essential to develop advanced automatic US image analysis methods to assist in US diagnosis and/or to make such assessment more objective and accurate. Deep learning has recently emerged as the leading machine learning tool in various research fields, and especially in general imaging analysis and computer vision. Deep learning also shows huge potential for various automatic US image analysis tasks. This review first briefly introduces several popular deep learning architectures, and then summarizes and thoroughly discusses their applications in various specific tasks in US image analysis, such as classification, detection, and segmentation. Finally, the open challenges and potential trends of the future application of deep learning in medical US image analysis are discussed.

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Ultrasound (US), as one of the most used imaging modalities, has been recognized as a powerful and ubiquitous screening and diagnostic tool for physicians and radiologists. In particular, US imaging is widely used in prenatal screening in most of the world due to its relative safety, low cost, noninvasive nature, real-time display, operator comfort, and operator experience [1]. Over the decades, it has been demonstrated that US has several major advantages over other medical imaging modalities such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT), including its non-ionizing radiation, portability, accessibility, and cost effectiveness. In current clinical practice, medical US has been applied to specialties such as echocardiography, breast US, abdominal US, transrectal US, intravascular US, and prenatal diagnosis US, which is specially used in obstetrics and gynecology (OB-GYN) [2]. However, US also presents unique challenges, such as low imaging quality caused by noise and artifacts, high dependence on abundant operator or diagnostician experience, and high inter- and intra-observer variability across different institutes and

manufacturers' US systems. For example, a study on the prenatal detection of malformations using US images demonstrated that the sensitivity ranged from 27.5% to 96% among different medical institutes [3]. To address these challenges, it is essential to develop advanced automatic US image analysis methods in order to make US diagnosis and/or assessment, as well as image-guided interventions/therapy, more objective, accurate, and intelligent.

Deep learning, which is a branch of machine learning, is considered to be a representation learning approach that can directly process and automatically learn mid-level and high-level abstract features acquired from raw data (e.g., US images). It holds the potential to perform automatic US image analysis tasks, such as lesion/nodule classification, organ segmentation, and object detection. Since AlexNet [4], a deep convolutional neural network (CNN) and a representative of the deep learning method, won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), deep learning began to attract attention in the field of machine learning. One year later, deep learning was selected as one of the top ten breakthrough technologies [5], which further consolidated its position as the leading machine learning tool in various research domains, and particularly in general imaging analysis (including natural and medical image analysis) and computer vision (CV). To date, deep learning has gained rapid development in terms of

* Corresponding authors.

E-mail addresses: nidong@szu.edu.cn (D. Ni), tfwang@szu.edu.cn (T. Wang).

<https://doi.org/10.1016/j.eng.2018.11.020>

2095-8099/© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

network architectures or models, such as deeper network architectures [6] and deep generative models [7]. Meanwhile, deep learning has been successfully applied to many research domains such as CV [8], natural language processing (NLP) [9], speech recognition [10], and medical image analysis [11–15], thus demonstrating that deep learning is a state-of-the-art tool for the performance of automatic analysis tasks, and that its use can lead to marked improvement in performance.

Recent applications of deep learning in medical US analysis have involved various tasks, such as traditional diagnosis tasks including classification, segmentation, detection, registration, biometric measurements, and quality assessment, as well as emerging tasks including image-guided interventions and therapy [16] (Fig. 1). Of these, classification, detection, and segmentation are the three most basic tasks. They are widely applied to different anatomical structures (organ or body location) in medical US analysis, such as breast [17,18], prostate [19–21], liver [22], heart/cardiac [23,24], brain [25,26], carotid [27,28], thyroid [29], intravascular [30,31], fetus [32–37], lymph node [38], kidney [39], spine [40], bone [41,42], muscle [43], nerve structure [44], tongue [45–47], and more. Multiple types of deep networks have involved these tasks. CNN is known as one of the most popular deep architectures and has also gained great success in various tasks, such as image classification [48,49], object detection [29,30], and target segmentation [44,50]. It is a common approach to apply a CNN model to learn from the obtained raw data (e.g., US images) in order to generate hierarchical abstract representations, followed by a softmax layer or other linear classifier (e.g., a support vector machine, SVM) that can be used to produce one or more probabilities or class labels. In this case, image annotations or labels are necessary for achieving the task. This is the so-called “supervised learning.” Unsupervised learning is also capable of learning representations from raw data [8,9]. Auto-encoders (AEs) and restricted Boltzmann’s machines (RBMs) are two of the most commonly applied unsupervised neural networks in medical US analysis promising improvements in performance. Unsupervised learning has one significant advantage

over supervised learning, which is that it does not require the utilization of time-consuming, labor-intensive, and expensive human annotations.

Although current medical US analysis still focuses on two-dimensional (2D) US image processing, there is a growing trend in applications of deep learning in three-dimensional (3D) medical US analysis. In the past two decades, commercial companies, together with researchers, have greatly advanced the development and progress of 3D US imaging technique. A 3D image (also commonly known as “3D volume”) is usually regarded as containing much richer information than a 2D image; thus, more robust results are attained when using a 3D volume as compared with a 2D image. More specifically, a 2D US image has certain inevitable limitations: ① Although US images are 2D, the anatomical structure is 3D; thus, the examiner/diagnostician must possess the ability to integrate multiple images in his or her mind (in an often inefficient and time-consuming process). Lack of this ability will lead to variability and incorrect diagnosis or misdiagnosis. ② Diagnostic (e.g., OB-GYN) and therapeutic (e.g., staging and planning) decisions often require accurate estimation of organ or tumor volume; however, 2D US techniques calculate volume from simple measurements of length, width, and height in two orthogonal views by assuming an idealized (e.g., ellipsoidal) shape. This may lead to low accuracy, high variability, and operator dependency. ③ A 2D US image presents a thin plane at various arbitrary angles in the body. These planes are difficult to localize and reproduce later for follow-up studies [51]. To overcome the limitations of 2D US, a variety of 3D US scanning, reconstruction, and display techniques have been developed, which provide a broad foundation for 3D medical US analysis. Furthermore, the current application of deep learning in medical US analysis is a growing trend that is supported by progress in this field [23,52].

Several review articles have been written to date on the application of deep learning to medical image analysis; these articles focus on either the whole field of medical image analysis [11–15] or other single-imaging modalities such as MRI [53] and microscopy [54].

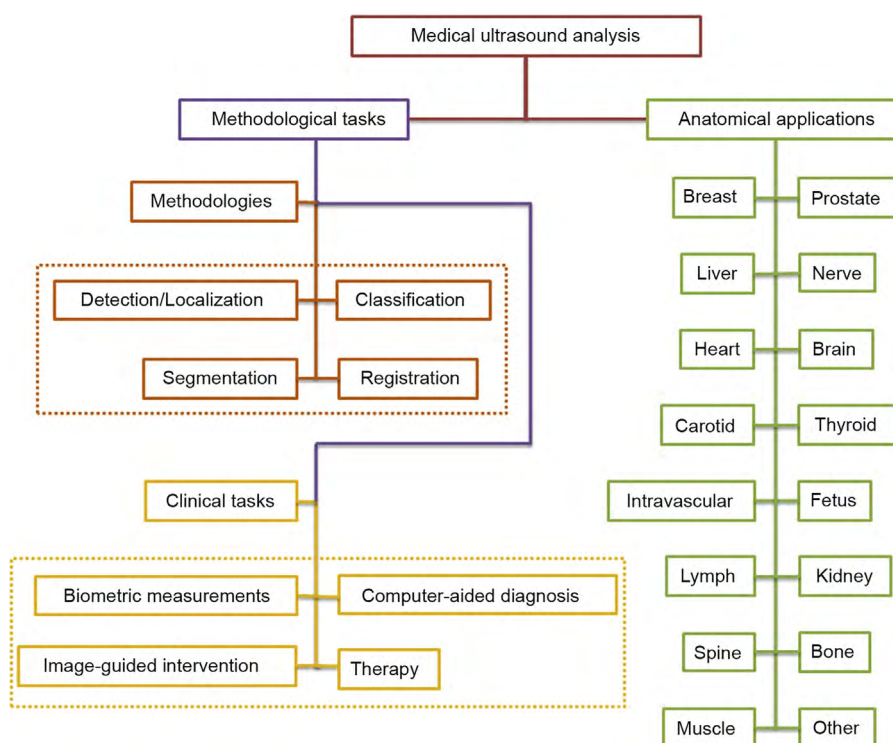


Fig. 1. Illustration of medical US analysis.

However, few focus on medical US analysis, aside from one or two papers that examine specific tasks such as breast US image segmentation [55]. A literature search for all works published in this field until 2018 Feb 1 was conducted by specifying key words (i.e., “ultrasound” OR “ultrasonography” OR “ultrasonic imaging” AND “convolutional” OR “deep learning”) in the main databases (e.g., PubMed and the Google Scholar database) and in several important conference proceedings (e.g., MICCAI, SPIE, ISBI, and EMBC). To screen the papers resulting from this search, the abstract of every paper was read in detail; papers that were relevant for this review were then chosen, which finally resulted in nearly 100 relevant papers, as summarized in Fig. 2 and Table S1 in Appendix A. This review attempts to offer a comprehensive and systemic overview of the use of deep learning in medical US analysis, based on typical tasks and their applications to different anatomical structures. The rest of the paper is organized as follows. In Section 2, we briefly introduce the basic theory and architectures of deep learning that are commonly applied in medical US analysis. In Section 3, we discuss in detail the applications of deep learning in medical US analysis, with a focus on traditional methodological tasks including classification, detection, and segmentation. Finally, in Section 4, we present potential future trends and directions in the application of deep learning in medical US analysis.

2. Deep learning architectures

Here, we start by briefly introducing the deep learning architectures that are widely applied in US analysis. Deep learning, as a branch of machine learning, essentially involves the computation of hierarchical features or representations of sample data (e.g., images), in which higher level abstract features are defined by combining them with lower level ones [9]. Based on the deep learning architectures and techniques in question, such as classification, segmentation, or detection, the deep learning architectures that are used in most of the current works in this field can be categorized into three major classes: ① supervised deep networks

or deep discriminative models, ② unsupervised deep networks or deep generative models, and ③ hybrid deep networks. The basic models or architectures applied in current medical US analysis are mainly CNNs, recurrent neural networks (RNNs), RBMs/DBNs (where DBN refers to deep belief networks), AEs, and variants of these deep learning architectures, as shown in Fig. 3. The term “hybrid” in the third category above refers to deep architecture that either comprises or makes use of both generative and discriminative model components, so that category is no longer specifically discussed here. Instead, we move on to introduce challenges and strategies in training the deep models that are commonly involved in medical US analysis. For convenience, several commonly used deep learning frameworks are also summarized in Section 2.4.

2.1. Supervised deep models

At present, supervised deep models are widely used for the classification, segmentation, and detection of anatomical structures in medical US images; for these tasks, CNNs and RNNs are the two most popular architectures. A brief overview of these two deep models follows.

2.1.1. Convolutional neural networks

CNNs are a type of discriminative deep architecture that includes several modules, each of which generally consists of a convolutional layer and a pooling layer. These are followed by other layers, such as a rectified linear unit (ReLU), and batch normalization if necessary. Fully connected layers generally follow in the last part of the network, to form a standard multi-layer neural network. In terms of structure, these modules are usually stacked, with one on top of another, to form a deep model; this makes it possible to take advantage of spatial and configuration information by taking in 2D or 3D images as input [8].

The convolutional layer shares many weights by performing convolution operations on input images. In fact, the role of a convolutional layer is to detect local features at different positions

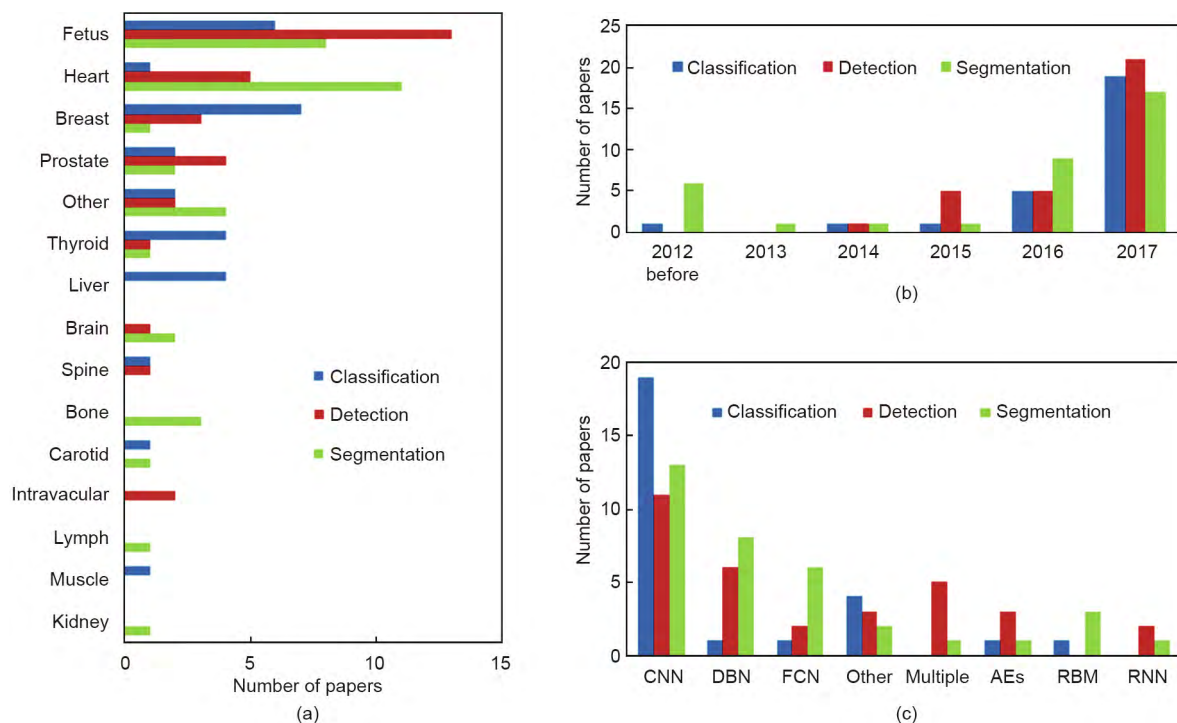


Fig. 2. Current applications of deep learning in medical US analysis. (a) Anatomical structures; (b) year of publication; (c) network architectures. DBN: deep belief network; FCN: fully convolutional network; Multiple: a hybrid of multiple network architectures; RNN: recurrent neural network; AEs include its variants, the sparse auto-encoder (SAE) and stacked denoising auto-encoder.

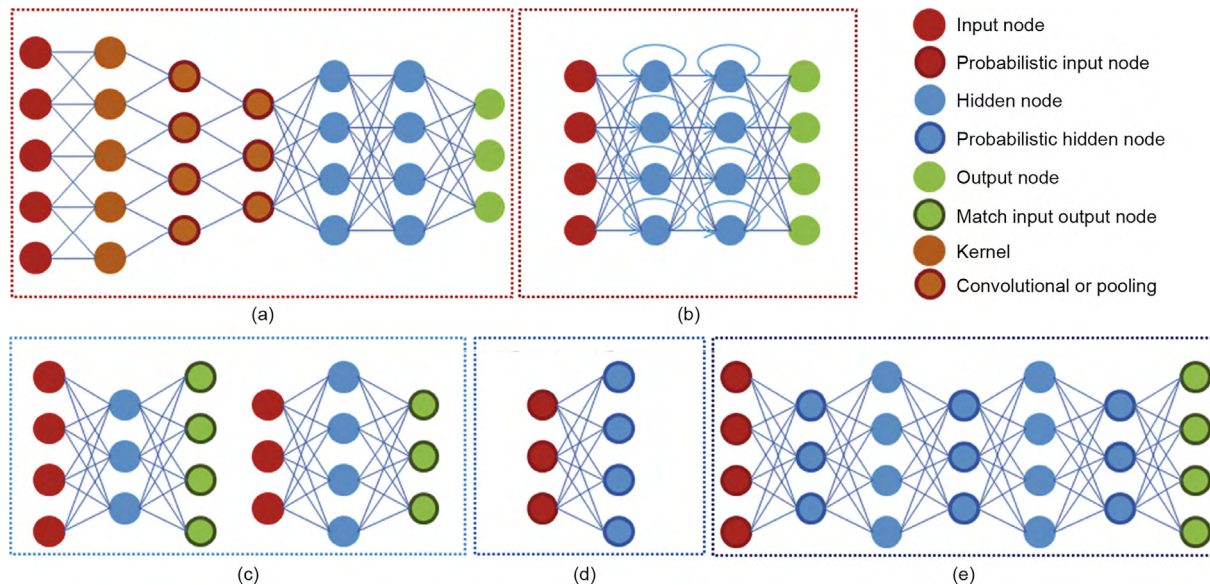


Fig. 3. Five representative neural network architectures, which can be categorized into two main types: ① supervised deep models, which include (a) CNNs and (b) RNNs; and ② unsupervised deep models, which include (c) AEs/SAEs (d) RBMs and (e) DBNs.

in the input feature maps (e.g., medical US images) with a set of k kernel weights $W = \{W_1, W_2, \dots, W_k\}$, together with the biases $b = \{b_1, b_2, \dots, b_k\}$, in order to generate a new feature map A_k^l . The convolutional process in every convolutional layer is expressed mathematically as follows:

$$A_k^l = \sigma(W_k^l * A^{l-1} + b_k^l) \quad (1)$$

where $\sigma(\cdot)$ is an element-wise nonlinear activation function, b_k^l is a bias parameter, and the asterisk, $*$, denotes a convolutional operator.

In a general CNN model, the determination of hyperparameters in a convolutional layer is crucial in order for the CNN to overcome reduction in the convolution process. This mainly involves three hyperparameters: depth, stride, and padding. The depth of the output volume corresponds to the number of filters, each of which learns to locally look for something different in the input. Specifying stride makes it possible to control how the filter convolves around the input volume. In practice, smaller strides always work better because small strides in the early layers of the network (i.e., those layers that are closer to the input data) can generate a large activation map, which can lead to better performance [56]. In a CNN with many convolutional layers, the reduction in output dimension can present a problem, since some regions—especially borders—are lost in every convolution operation. Padding (generally zero-padding) around the border in the input volume is one of the strategies that is most commonly used to eliminate the effect of dimensional reduction in the convolution process. One of the greatest benefits of padding is that it makes it possible to design deeper networks. In addition, padding actually improves performance because it prevents information loss at the borders in the input volume. That is, under the conditions of limited computational cost and time cost, it is necessary to perform trade-offs between multiple factors (i.e., the number of filters, filter size, strides and network depth, etc.) for a specific task in practice.

The output of the convolutional layer is subsampled by the subsequent pooling layer in order to reduce the data rate from the layer below. Together with appropriately chosen pooling schemes, the weight shared in the convolutional layer can imbue the CNN with certain invariant properties such as translational invariance. This can also greatly reduce the number of parameters; for exam-

ple, the number of weights no longer absolutely depends on the size of the input images. Note that fully connected layers, which are generally added at the end of the convolutional stream of the network, usually no longer share the weights. In a standard CNN model, a distribution over classes is generally achieved by feeding the activations through a softmax function in the last layer of the network; however, several conventional machine learning methods use an alternative, such as voting strategy [57] or linear SVM [58].

Given their increasing popularity and practicability, many classical and CNN-based deep learning architectures have been developed and applied in (medical) image analysis, NLP, and speech recognition. Examples include AlexNet (or CaffeNet, which is suitable for the Caffe deep learning framework), LeNet, faster R-CNN, GoogLeNet, ResNet, and VGGNet; please refer to Ref. [59] for a detailed comparison of these architectures in terms of various performance indicators (e.g., accuracy, inference time, memory, and parameters utilization).

2.1.2. Recurrent neural networks

In practical terms, an RNN is generally considered to be a type of supervised deep network that is used for a variety of tasks in medical US analysis [21,60]. In an RNN, the depth of the network can be as long as the length of the input sample data sequences (e.g., medical US video sequences). A plain RNN contains a latent or hidden state, h_t , at time t that is the output of a nonlinear mapping from its input, x_t , and the previous state h_{t-1} , expressed as follows:

$$h_t = \sigma(Wx_t + Rh_{t-1} + b) \quad (2)$$

where the weights W and R are shared over time, b is a bias parameter.

The RNN has an inherent advantage for modeling sequence data (e.g., medical US video sequences) due to the structural characteristics of the network. However, until recently, RNNs have not been widely utilized in the various study tasks that are referred to as sequence models. This is partly because it is difficult to train the RNN itself to capture long-term dependencies, in which the RNN usually gives rise to gradient explosion or gradient vanishing problems that were discovered in the early 1990s [61]. Therefore, several specialized memory units have been developed, the earliest

and most popular of which are long short-term memory (LSTM) cells [62] and their simplification-gated recurrent unit [63]. Thus far, RNNs are mainly applied in speech or text-recognition domains, and are rarely used in medical image analysis, much less medical US analysis.

RNN can also be considered to be a type of deep model for unsupervised learning. In the unsupervised learning mode, the RNN is usually used to predict the subsequent data sequences using the previous data samples. It does not need additional class information (e.g., target class labels) to help learning, although a label sequence is essential for learning in the supervised mode.

2.2. Unsupervised deep models

Unsupervised learning means that task-specific supervision information (e.g., the annotated target class labels) is unnecessary in the learning process. In practice, various deep models with unsupervised learning are utilized to generate data samples by sampling from the networks, such as AE, RBMs/DBNs, and generalized denoising AE [64]. From this perspective, unsupervised deep models are usually regarded as generative models to be applied in a variety of tasks. Below, we briefly introduce the three basic deep models for unsupervised feature/representation learning that are used most in medical US analysis.

2.2.1. The auto-encoder and its variants

Simply speaking, the AE is a nonlinear feature-extraction approach that does not involve the use of target class labels. This approach is usually used for representation learning or for effective encoding of the original input data (e.g., in the form of input vectors) in hidden layers [9]. As such, the extracted features are focused on conserving and better representing information, rather than on performing specific tasks (e.g., classification), although these two goals are not always mutually exclusive.

An AE is typically a simple network that includes at least three layers: an input layer, x , which represents the original data or input feature vectors (e.g., patches/pixels in an image or spectrum in a speech) one or more hidden layers, h , which denote the transformed features; and an output layer, y , which matches the input layer x for reconstruction through the nonlinear function σ in order to activate the hidden layers:

$$h = \sigma(Wx + b) \quad (3)$$

To date, many variants of AEs have been developed. Examples include sparse auto-encoders (SAEs) [64] and denoising auto-encoders (DAEs) and their stacked versions [65]. In an SAE model, regularization and sparsity constraints are adopted in order to enhance the solving process in the training network, while “denoising” is used as a solution to prevent the network from learning a trivial solution. The stacked version of these models is usually generated by placing the AE layers on top of each other.

2.2.2. Restricted Boltzmann's machines and deep belief networks

An RBM is a particular type of Markov's random field with a two-layer architecture [66]. In terms of structure, it is a single-layer undirected graphical model consisting of a visible layer and a hidden layer, with symmetric connectivity between them and no connectivity among units within the same layer. Therefore, it is essentially an AE [67]. In practice, an RBM is rarely used alone; rather, it is stacked one by one to generate a deeper network, which typically results in a single probabilistic model called a DBN.

A DBN consists of a visible layer and several hidden layers; the top two layers form an undirected bipartite graph (e.g., an RBM) and the lower layers form a sigmoid belief network with directed

and top-down connections. A DBN is capable of good generalization because it can be pre-trained layer-wise using unlabeled data; this is practically accomplished using a small number of labeled training samples. Since the DBN is trained in an unsupervised manner, a final fine-tuning step is necessary for a specific task in practice; this is done by providing a supervised optimization option by adding a linear classifier (e.g., SVM) to the top layer of the DBN. For unsupervised learning models, a fine-tuning step that follows after the final representation learning is also a practical and common solution to address a specific task such as image classification, object detection, or organ segmentation.

2.3. Challenges and strategies in training deep models

The great success of deep learning comes from the fact that a large number of labeled training samples are required in order to achieve excellent learning performance. However, this requirement is difficult to meet in current medical US analysis, where expert annotation is expensive and where some diseases (e.g., lesions or nodules) are scarce in the datasets [68]. Therefore, the question of how to train a deep model using a limited training sample has become an open challenge in medical US analysis. One of the most common problems when using limited training samples is that it is easy to over-fit the deep model. To address the issue of model overfitting, two main pathways can be selected: model optimization and transfer learning. For model optimization, several fruitful strategies such as well-designed initialization strategies, stochastic gradient descent and its variants (e.g., momentum and Adagrad [69]), efficient activation functions, and other powerful intermediate regularization strategies (e.g., batch normalization) have been proposed and constantly improved in recent years, as follows [11]:

(1) Well-designed initialization/momentum strategies [70] refer to the utilization of well-designed random initialization and a particular type of schedule in order to slowly increase the momentum parameter on the iterations of the training model.

(2) Efficient activation functions, such as ReLU [71,72], perform a nonlinear operation generally following the convolutional layer. In addition, Maxout [73] is actually a type of activation function being particularly suited for training with dropout.

(3) Dropout [74] randomly deactivates the units/neurons in a network at a certain rate (e.g., 0.5) on each training iteration.

(4) Batch normalization [75] performs the normalization operation for each training mini-batch and back-propagates the gradients through the normalized parameters on each training iteration.

(5) Stack/denoising [65] is mainly used for AEs in order to make the model deeper and reconstruct the original “clean” inputs from the corrupted ones.

Another key solution is transfer learning, which has also been widely adopted and which exhibits excellent capacity for improving the performance of learning without the need for large samples. This method avoids expensive data-labeling efforts in the application of a specific domain. According to Pan and Yang [76], transfer learning is categorized into three settings: *inductive transfer learning*, in which the target and the source tasks are different, regardless of whether the target and source domains are the same or not; *transductive transfer learning*, in which the target task is the same as the source task, while the target domains are different from the source domains; and *unsupervised transfer learning*, which is similar to inductive transfer learning, except that the target task differs from but is related to the source task. Based on what is being transferred, the approaches used for the abovementioned three different settings of transfer learning can be classified into four cases: the instance approach, the representation approach, the parameter-transfer approach, and the relational knowledge approach. However, this review is most concerned with how to

improve the performance by transferring knowledge from another domains (in which it is easy to collect a large number of training samples, e.g., CV, speech, and text) to the medical US domain. This process involves two main strategies: ① using a pre-trained network as a feature extractor (i.e., to learn features from scratch); and ② fine-tuning a pre-trained network on medical US images or video sequences—a method that is widely applied at present in US analysis. Both strategies achieve excellent performance in several specific tasks [77,78].

Some additional strategies need to be noted, such as data pre-processing and data augmentation/enhancement [4,16].

2.4. Popular deep learning frameworks

With the rapid development of the relevant hardware (e.g., graphics processing unit, GPU) and software (e.g., open-source software libraries), deep learning techniques have quickly become popular in various research domains throughout the world. Five of the most popular open-source deep learning frameworks (i.e., packages) are listed below:

- (1) Caffe [79]: <https://github.com/BVLC/caffe>;
- (2) Tensorflow [80]: <https://github.com/tensorflow/tensorflow>;
- (3) Theano [81]: <https://github.com/Theano/Theano>;
- (4) Torch7/PyTorch [82]: <https://github.com/torch/torch7> or <https://github.com/pytorch/pytorch>; and
- (5) MXNet [83]: <https://github.com/apache/incubator-mxnet>.

Most of the popular frameworks provide multiple interfaces, such as C/C++, MATLAB, and Python. In addition, several packages provide a higher level library written on top of these frameworks, such as Keras[†]. For the advantages and disadvantages of these frameworks, please refer to Ref. [84]. In practice, researchers can choose any framework, or use their own written frameworks, based on the actual requirements and personal preferences.

3. Applications of deep learning in medical US analysis

As noted earlier, current applications of deep learning techniques in US analysis mainly involve three types of tasks: classification, detection, and segmentation for various anatomical structures or tissues, such as the breast, prostate, liver, heart, and fetus. In this review, we discuss the application of each task separately for various anatomical structures. Furthermore, 3D US presents a promising trend in improving US imaging diagnosis in clinical practice, which is discussed in detail as a separate sub-section.

3.1. Classification

The classification of images is a fundamental cognitive task in diagnostic radiology, which is accomplished by the identification of certain anatomical or pathological features that can discriminate one anatomical structure or tissue from others. Although computers are currently far from being able to reproduce the full chain of reasoning required for medical image interpretation, the automatic classification of targets of interest (e.g., tumors/lesions, nodules, fetuses) is a research focus in computer-aided diagnosis systems. Traditional machine learning methods often utilized various hand-crafted features extracted from US images in combination with a multi-way linear classifier (e.g., SVM) in order to achieve a specific classification task. However, these methods are susceptible to image distortion, such as deformation due to the internal or external environments, or to conditions in the imaging process. Here, deep neural networks (DNNs) have several obvious advantages due to their direct learning of mid- and high-level abstract features

from the raw data (or images). In addition, DNNs can be directly used to output an individual prediction label for each image, in order to classify targets of interest. For different anatomical application areas, several unique challenges exist, which are discussed below.

3.1.1. Tumors or lesions

According to the latest statistics from the Centers for Disease Control and Prevention[‡], breast cancer has become the most common cancer and the second leading cause of cancer death among women around the world. Although mammography is still the leading imaging modality for screening or diagnosis in clinical practice, US imaging is also a vital screening tool for the diagnosis of breast cancer. In particular, the use of US-based computer-aided diagnosis (CADx) for the classification of tumor diseases provides an effective decision-making support and a second tool option for radiologists or diagnosticians. In a conventional CADx system, feature extraction is the foundation on which subsequent steps, including feature selection and classification, can be integrated in order to achieve the final classification of tumors or mass lesions. Traditional machine learning approaches for breast tumor or mass lesion CADx often utilize handcrafted and heuristic lesion-extracted features [85]. In contrast, deep learning can directly learn features from images in an automatic manner.

As early as 2012, Jamieson et al. [86] performed a preliminary study that referred to the use of deep learning in the task of classifying breast tumors or mass lesions. As illustrated in Fig. 4(a), adaptive deconvolutional networks (ADNs), which are unsupervised and generative hierarchical deep models, were utilized to learn image features from diagnostic breast tumor or mass lesion US images and generate feature maps. Post-processing steps that include building image descriptors and a spatial pyramid matching (SPM) algorithm are performed. Because the model was trained in an unsupervised fashion, the learned high-level features (e.g., the SPM kernel output) were regarded as the input to train a supervised classifier (e.g., a linear SVM), in order to achieve binary classification between malignant and benign breast mass lesions. The results showed that the performance reached the level of conventional CADx schemes that employ human-designed features. Following this success, many similar studies applied deep learning methods to breast tumor diagnosis. Both Liu et al. [87] and Shi et al. [19] employed a supervised deep learning algorithm called a deep polynomial network (DPN), or its stacked version, namely stacked DPN (S-DPN), on two small US datasets. With the help of preprocessing (i.e., the shearlet transform based texture feature extraction and region of interest (ROI) extraction) and a SVM classifier (or multiple kernel learning), the highest classification accuracies of 92.4% is obtained, outperforming the unsupervised deep learning algorithms, such as stacked AE and DBN. This approach is a good alternative solution to the problem of a local patch being unable to provide rich contextual information when using deep learning to learn image representation from patch-level US samples. In addition, the stacked denoising auto-encoder (SDAE) [88], a combination of the point-wise gated Boltzmann machine (PGBM) and RBM [89], and the GoogLeNet CNN [90] have been applied to breast US or shear-wave elastography images for breast cancer diagnosis; all of these obtained a superior performance when compared with human experts. In the work of Antropova et al. [91], a method that fuses the extracted and pooled low- and mid-level features using a pre-trained CNN with hand-designed features using conventional CADx methods was applied to three clinical imaging modality datasets, and demonstrated significant performance improvement.

[†] <https://github.com/fchollet/keras>.

[‡] <https://www.cdc.gov/cancer/dccp/data/women.htm>.

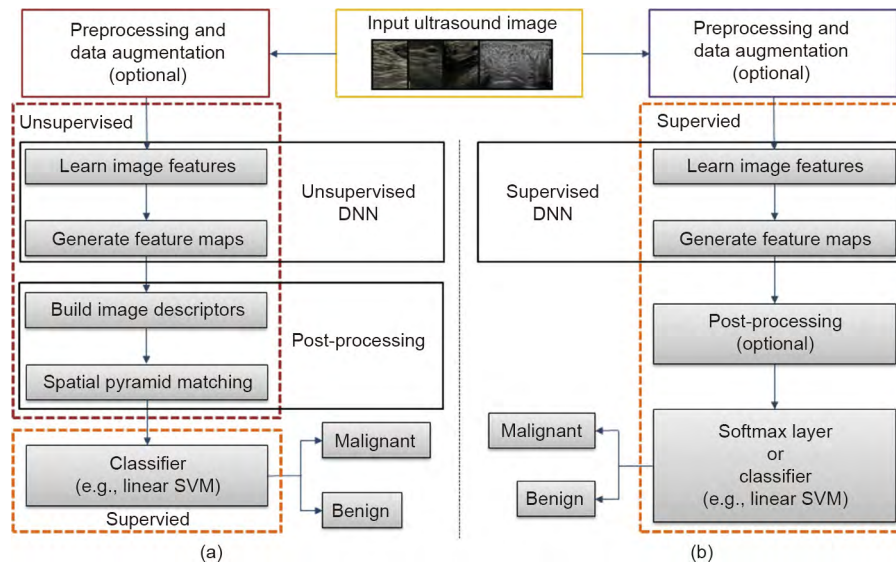


Fig. 4. Flowcharts of (a) unsupervised deep learning and (b) supervised deep learning for tumor US image classification. It is usually optional to perform the preprocessing and data-augmentation steps (e.g., ROI extraction, image cropping, etc.) for US images before using them as inputs to deep neural networks. Although the post-processing step also applies to supervised deep learning, few researchers do this; instead, the feature maps are directly used as inputs to a softmax classifier for classification.

Another common tumor is liver cancer, which has become the sixth most common cancer and the third leading cause of cancer death worldwide [92]. Early accurate diagnosis is very important in increasing survival rates by providing optimal interventions. Biopsy is still the current golden standard for liver cancer diagnosis, and is heavily relied upon by conventional CADx methods. However, biopsy is invasive and uncomfortable, and can easily cause other adverse effects. Therefore, US-based diagnostic techniques have become one of the most important noninvasive methods for the detection, diagnosis, intervention, and treatment of liver cancer. Wu et al. [22] applied a three-layer DBN in time-intensity curves (TICs) extracted from contrast-enhanced US (CEUS) video sequences in order to classify malignant and benign focal liver lesions. They achieved a highest accuracy of 86.36%, thus outperforming conventional machine methods such as linear discriminant analysis (LDA), k -nearest neighbors (k -NN), SVM, and back propagation net (BPN). To reduce computational complexity using TIC-based feature-extraction methods, Guo et al. [93] adopted deep canonical correlation analysis (DCCA)—a variant of canonical correlation analysis (CCA)—combined with a multiple kernel learning (MKL) classifier—a typical multi-view learning approach—in order to distinguish benign liver tumors from malignant liver cancers. They demonstrated that taking full advantage of these two methods can result in high classification accuracy (90.41%) with low computational complexity. In addition, the transfer learning strategy is frequently adopted for liver cancer US diagnosis [58,94].

3.1.2. Nodules

Thyroid nodules have become one of the most common nodular lesions in the adult population worldwide. At present, the diagnosis of thyroid nodules relies on non-surgical (mainly fine needle aspiration (FNA) biopsy) and surgical (i.e., excisional biopsy) methods. However, both of these methods are too labor-intensive for large-scale screenings, and may make the patients anxious and increase the costs. With the rapid development of US techniques, US has become an alternative tool for the diagnosis and follow-up of thyroid nodules due to its real-time and noninvasive nature. To alleviate operator dependence and improve diagnostic performance, US-based CADx systems have been developed to detect

and classify thyroid nodules. Ma et al. [95] integrated two pre-trained CNNs in a fusion framework for thyroid nodule diagnosis: One was a shallower network that was preferable for learning low-level features, and the other was a deeper network that was good at learning high-level abstract features. More specifically, the two CNNs were trained on a large thyroid nodule US dataset separately, and then the two learned feature maps were fused as input into a softmax layer in order to diagnose thyroid nodules. Integrating the learned high-level features from CNNs and conventional hand-designed low-level features is an alternative scheme that is demonstrated in Liu et al. [96,97]. In order to overcome the problem of redundancies and irrelevancies in the integrated feature vectors, and to avoid overfitting, it is necessary to select a feature subset. The results indicated that this method can improve the accuracy by 14%, compared with the traditional features. In addition, efficient preprocessing and data-augmentation strategies for a specific task have been demonstrated to improve the diagnosis performance [48].

3.1.3. Fetuses and neonates

In prenatal US diagnosis, fetal biometry is an examination that includes an estimation of abdominal circumference (AC); however, it is more difficult to perform an accurate measurement of AC than of other parameters, due to low and non-uniform contrast and irregular shape. In clinical examination and diagnosis, incorrect fetal AC measurement may lead to inaccurate fetal weight estimation and further increase the risk of misdiagnosis [98]. Therefore, quality control for fetal US imaging is of great importance. Recently, Wu et al. [99] proposed a fetal US image quality assessment (FUIQA) scheme with two steps: ① A CNN was used to localize the ROI, and ② based on the ROI, another CNN was employed to classify the fetal abdominal standard plane. To improve the performance, the authors adopted several data-enhancement strategies such as local phase analysis and image cropping. Similarly, Jang et al. [100] employed a specially designed CNN architecture to classify image patches from an US image into the key anatomical structures; based on the accepted fetal abdominal plane (i.e., the standard plane), fetal AC measurement was then estimated through an ellipse detection method based on the Hough transform. Gao et al. [101] explored the transferability of features

learned from large-scale natural images to small US images through the multi-label classification of fetal anatomical structures. The results demonstrated that transferred CNNs outperformed those that were directly learned from small US data (91.5% vs. 87.9%).

The location of the fetal heart and classification of the cardiac view are very important in aiding the identification of congenital heart diseases. However, these are challenging tasks in clinical practice due to the small size of the fetal heart. To address these issues, Sundaresan et al. [102] posed the solution as a semantic segmentation problem. More specifically, a fully convolutional neural network (FCN) was applied to segment the fetal heart views from the US frames, allowing the detection of the heart and classification of the cardiac views to be accomplished in a single step. Several post-processing steps were adopted to address the problem of the predicted image possibly including multiple labels of regions of different non-background. In addition, Perrin et al. [103] directly trained a CNN on echocardiographic images/frames from five different pediatric populations to differentiate between congenital heart diseases. In a specific fetal standard plane recognition task, a very deep CNN with a global average pooling (GAP) strategy achieved significant performance improvement on the limited training data [104,105].

3.2. Detection

The detection of objects of interest (e.g., tumors, lesions, and nodules) on US images or video sequences is essential in US analysis. In particular, tumor or lesion detection can provide strong support for object segmentation and for differentiation between malignant and benign tumors. Anatomical object (e.g., fetal standard plane, organs, tissues, or landmarks) localization has also been regarded as a prerequisite step for segmentation tasks or clinical diagnosis workflow for image-based intervention and therapy.

3.2.1. Tumors or lesions

Detection or localization of tumors/lesions is vital in the clinical workflow for therapy planning and intervention, and is also one of the most labor-intensive tasks. There are several overt differences in the detection of lesions in different anatomical structures. This task typically consists of the localization and identification of small lesions in the full image space. Recently, Azizi et al. [20,106,107] successfully accomplished the detection and grading of prostate cancer through a combination of high-level abstract features extracted from temporal-enhanced US by a DBN and the structure of tissue from digital pathology. To perform a comprehensive comparison, Yap et al. [108] contrasted three different deep learning methods—a patch-based LeNet, a U-net, and a transfer learning approach with a pre-trained FCN-AlexNet—for breast lesion detection from two US image datasets acquired from two different US systems. Experiments on two breast US image datasets indicated that an overall detection performance improvement was obtained by the deep learning algorithms; however, no single deep model could achieve the best performance in terms of true positive fraction (TPF), false positive per image (FPs/image), and *F*-measure. Similarly, Cao et al. [109] performed a comprehensive comparison among four state-of-the-art CNN-based object detection deep models—Fast R-CNN [110], Faster R-CNN [111], You Only Look Once (YOLO) [112], and Single-Shot MultiBox Detector (SSD) [113]—for breast lesions detection, and demonstrated that SSD achieved the best performance in terms of both precision and recall.

3.2.2. Fetus

As a routine obstetric examination for all pregnant women, fetal US screening plays a critical role in confirming fetal viability,

establishing gestational age accurately, and looking for malformation that could influence prenatal management. Among the workflow of fetal US diagnosis, acquisition of the standard plane is the prerequisite step and is crucial for subsequent biometric measurements and diagnosis [114]. In addition to the use of traditional machine learning methods for the detection of the fetal US standard plane [115,116], there has recently been an increasing trend in the use of deep learning algorithms to detect the fetal standard plane. Baumgartner et al. [117,118] and Chen et al. [78,119] accomplished the detection of 13 fetal standard views (e.g., kidneys, brain, abdominal, spine, femur, and cardiac plane) and the fetal abdominal (or face and four-chamber view) standard plane in 2D US images through the transferred deep models, respectively. To incorporate the spatiotemporal information, a transferred RNN-based deep model has also been employed for the automatic detection of multiple fetal US standard planes (e.g., abdominal, face axial, and four-chamber view) in US videos [60]. Furthermore, Chen et al. [120] presented a general framework based on a composite framework of the convolutional and RNNs for the detection of different standard planes from US videos.

3.2.3. Cardiac

Accurate identification of cardiac cycle phases (end-diastolic (ED) and end-systolic (ES)) in echocardiograms is an essential prerequisite for the estimation of several cardiac parameters such as stroke volume, ejection fraction, and end-diastolic volume. Dezaki et al. [121] proposed a deep residual recurrent neural network (RRN) to automatically recognize cardiac cycle phases. RRNs comprise residual neural networks (ResNet), two blocks of LSTM units, and a fully connected layer, and thus combines the advantage of the ResNet, which handles the vanishing or exploding gradient problem when the CNN goes deeper, and that of the RNN (LSTM), which is able to model the temporal dependencies between sequential frames. Similarly, Sofka et al. [122] presented a fully convolutional regression network for the detection of measurement points in the parasternal long-axis view of the heart; this network contained an FCN to regress the point locations and LSTM cells to refine the estimated point location. Note that reinforcement learning has also been combined with deep learning for anatomical (cardiac US) landmark detection [123].

3.3. Segmentation

The segmentation of anatomical structures and lesions is a prerequisite for the quantitative analysis of clinical parameters related to volume and shape in cardiac or brain analysis. It also plays a vital role in detecting and classifying lesions (e.g., breast, prostate, thyroid nodules, and lymph node) and in generating ROI for subsequent analysis in a CADx system. Accurate segmentation of most anatomical structures, and particularly of lesions (nodules) in US images, is still a challenging task due to low contrast between the target and background in US images. Furthermore, it is well known that manual segmentation methods are time consuming and tedious, and suffer from great individual variability. Therefore, it is imperative to develop more advanced automatic segmentation methods to solve these problems. Examples of some results of anatomical structure segmentation using deep learning are illustrated in Fig. 5 [21,38,44,46,50,57,124–126].

3.3.1. Non-rigid organs

Echocardiography has become one of the most commonly used imaging modalities for visualizing and diagnosing the left ventricle (LV) of the heart due to its low cost, availability, and portability. In order to diagnose cardiopathy, a quantitatively functional analysis of the heart must be done by a cardiologist, which is often based on accurate segmentation of the LV at the end-systole and

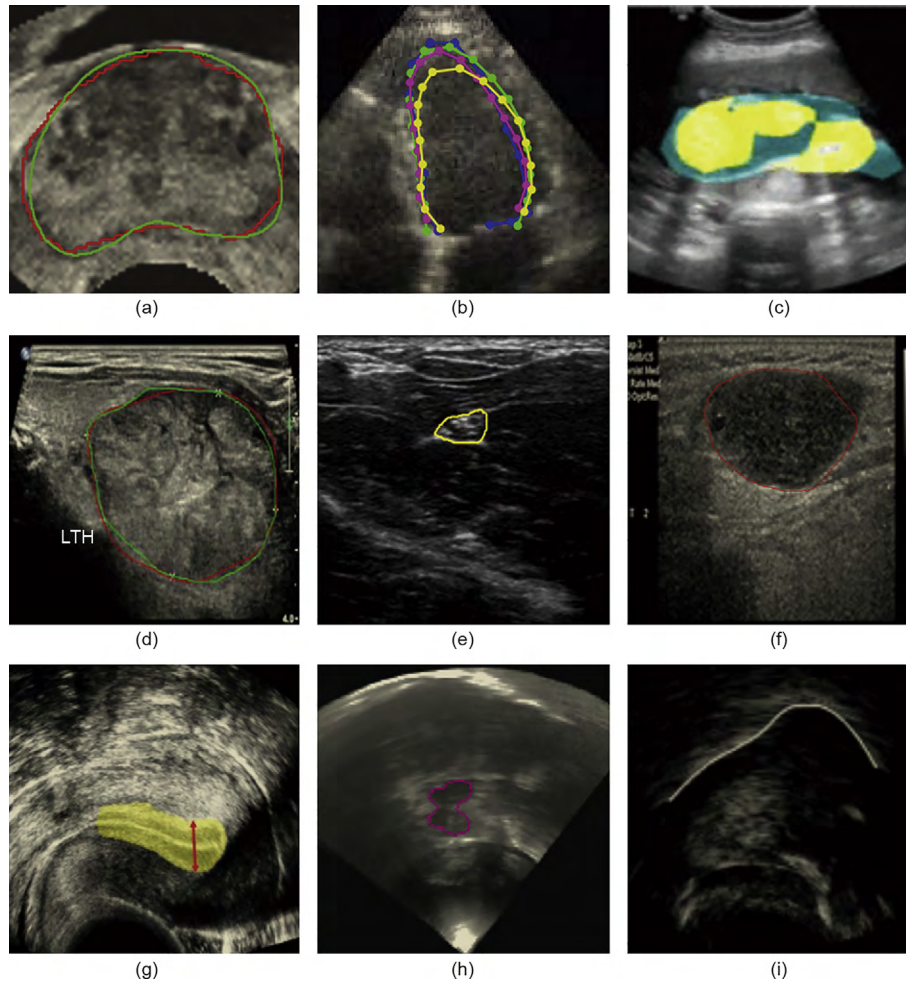


Fig. 5. Examples of segmentation results from certain anatomical structures using deep learning. (a) prostate [21]; (b) left ventricle of the heart [124]; (c) amniotic fluid and fetal body [50]; (d) thyroid nodule [125]; (e) median nerve structure [44]; (f) lymph node [38]; (g) endometrium [126]; (h) midbrain [57]; (i) tongue contour [46]. All of these results demonstrated a segmentation performance that was comparable with that of human radiologists. Lines or masks of different colors represent the corresponding segmented contours or regions.

end-diastole phases. It is obvious that manual segmentation of the LV is tedious, time consuming, and subjective, problems that can potentially be addressed by an automatic LV segmentation system. However, fully automatic LV segmentation is a challenging task due to significant appearance and shape variations, a low signal-to-noise ratio, shadows, and edge dropout. To address these issues, various conventional machine learning methods such as active contours [127] and deformable templates [128] have been widely used to successfully segment the LV, under the condition of using prior knowledge about the LV shape and appearance. Recently, deep learning-based methods have also been frequently adopted. Carneiro et al. [129–134] employed DNNs that are capable of learning high-level features from the original US images to automatically segment the LV of the heart. To improve the segmentation performance, several strategies (e.g., efficient search methods, particle filters, an online co-training method, and multiple dynamic models) were also adopted.

The typical non-rigid segmentation approach often divides the segmentation problem into two steps: ① rigid detection and ② non-rigid segmentation or delineation. The first step is of great importance because it can reduce the search time and training complexities. To reduce the complexity of training and inference in a rigid detection while maintaining the segmentation accuracy, Nascimento and Carneiro [124,135] utilized a sparse manifold learning method combined with DBN to segment non-rigid objects.

Their experiments demonstrated that the combination of sparse manifold learning and DBN in the rigid detection stage yielded a performance as accurate as the state of the art, but with lower training and search complexity. Unlike the typical non-rigid segmentation scheme, Nascimento and Carneiro [136] directly performed non-rigid segmentation through the sparse low-dimensional manifold mapping of explicit contours, but with a limited generalization capability. Although most studies have demonstrated that the use of deep learning can yield a much superior performance when compared with conventional machine learning methods, a recent study [137] showed that handcrafted features outperformed the CNN on LV segmentation in 2D echocardiographic images at a markedly lower computational cost in training. A plausible explanation is that the supervised descent method (SDM) [138] regression method applied to hand-designed features is more flexible in iteratively refining the estimated LV contour.

Compared with adult LV segmentation, fetal LV segmentation is more challenging, since fetal echocardiographic sequences suffer from inhomogeneities, artifacts, poor contrast, and large inter-subject variations; furthermore, there is usually a connected LV and left atrium (LA) due to the random movements of the fetus in the womb. To tackle these problems, Yu et al. [139] proposed a dynamic CNN method based on multiscale information and fine-tuning for fetal LV segmentation. The dynamic CNN was fine-tuned by deep tuning with the first frame and shallow tuning

with the rest of frames in each echocardiographic sequence, respectively, in order to adapt to the individual fetus. Furthermore, a matching method was utilized to separate the connection area between the LV and LA. The experiments showed that the dynamic CNN obtained a remarkable performance improvement from 88.35% to 94.5% in terms of the mean of the Dice coefficient, when compared with the fixed CNN.

3.3.2. Rigid organs

Boundary incompleteness is a common problem for many anatomical structures (e.g., prostate, breast, kidney, fetus, etc.) in medical US images, and presents great challenges to the automatic segmentation of these structures. Two main methodologies are currently used to address this issue: ① a bottom-up manner that classifies each pixel into foreground (object) or background in a supervised manner; and ② a top-down manner that takes advantage of prior shape information to guide segmentation. By classifying each pixel in an image in an end-to-end and fully supervised learning manner, many studies reached the task of pixel-level segmentation for different anatomical structures, such as fetal body and amniotic fluid [50], lymph node [38], and bone [140]; all of the deep learning methods presented in these studies outperformed the state-of-the-art methods in both performance and speed in terms of the specific task.

One significant advantage of the bottom-up manner is that it can provide a prediction for each pixel in an image; however, it may be unable to deal with boundary information loss due to a lack of the prior shape information. In contrast, the top-down manner can provide strong shape guidance for the segmentation task by modeling the shape, although appropriate shape modeling is difficult. In order to simultaneously accomplish landmark descriptor learning and shape inference, Yang et al. [21] formulated boundary completeness as a sequential problem—namely, modeling for shape in a dynamic manner. To take advantage of both the bottom-up and top-down methods, Ravishankar et al. [39] employed a shape that had been previously learned from a shape regularization network to refine the predicted segmentation result obtained from an FCN segmentation network. The results on a kidney US image dataset demonstrated that incorporation of the prior shape information led to an improvement of approximately 5% in the kidney segmentation. In addition, Wu et al. [141] implanted the FCN core into an auto-context scheme [142] in order to take advantage of local contextual information, and thus bridge severe boundary incompleteness and remarkably improve the segmentation accuracy. Anas et al. [143] applied an exponential weight map in the optimization of a ResNet-based deep framework to improve the local prediction.

Another way to perform the segmentation task is to formulate the problem as a patch-level classification task, as was done in Ref. [125]. This method can significantly reduce the extensive computation cost and memory requirement.

3.4. 3D US analysis

Due to the difficulty of 3D deep learning, the deep learning methods that are currently applied in medical US analysis mostly work on 2D images, although the input may be 3D. In fact, 3D deep learning is still a challenging task, due to the following limitations: ① Training a deep network on a large volume may be too computationally expensive (e.g., with a significantly increased memory and computational requirement) for real clinical application; and ② a deep network with a 3D patch as input require more training samples, since a 3D network contains parameters that are orders of magnitude higher than a 2D network. This may dramatically increase the risk of overfitting, given the limited training data [144]. In contrast, US image analysis fields often struggle with

limited training samples (usually in the hundreds or thousands, even after using data-augmentation strategies) due to the difficulty in generating and sharing lesion or disease images. Nevertheless, in the domain of medical US analysis, an increasing number of attempts are being made to address these challenging 3D deep learning tasks.

In routine gynecological US examination and in endometrium cancer screening in women with post-menopausal bleeding, endometrium assessment via thickness measurement is commonly performed. Singhal et al. [126] presented a two-step algorithm based on FCN to accomplish the fully automated measurement of endometrium thickness. First, a hybrid variational curve-propagation model, called the deep-learned snake (DLS) segmentation model, was presented in order to detect and segment the endometrium from 3D transvaginal US volumes. This model integrated a deep-learned endometrium probability map into the segmentation energy function, with the map being predictively built on U-net-based endometrium localization in a sagittal slice. Following the segmentation, the thickness was measured as the maximum distance between the two interfaces (basal layers) in the segmented mask.

To address the problem of automatic localization of the needle target for US-guided epidural needle injections in obstetrics and chronic pain treatment, Pesteie et al. [145] proposed a convolutional network architecture along with a feature-augmentation technique. This method has two steps: ① plane classification using local directional Hadamard (LDH) features and a feed-forward neural network from 3D US volumes; and ② target localization by classifying pixels in the image via a deep CNN within the identified target planes.

Nie et al. [146] proposed a method for automatically detecting the mid-sagittal plane based on complex 3D US data. To avoid unnecessary massive searching and the corresponding huge computation load, they subtly turned the sagittal plane detection problem into a symmetry plane and axis searching problem. More specifically, the proposed method consisted of three steps: ① A DBN was built to detect an image patch fully containing the fetal head from the middle slice of the 3D US data, as proposed in Ref. [147]; ② an enhanced circle-detection method was used to localize the position and size of the fetal head in the image patch; and ③ finally, the sagittal plane was determined by a model, with prior knowledge of the position and size of the fetal head having been obtained in the first two steps.

It should be pointed out that all three methods are actually 2D deep learning-based approaches in a slice-by-slice fashion, although both can be used in 3D US volumes. Here, the advantage is high speed, low memory consumption, and the ability to utilize pre-trained networks either directly or via transfer learning. However, the disadvantage is being unable to exploit the anatomical contextual information in directions orthogonal to the image plane. To address this disadvantage, Milletari et al. [57] proposed a patch-wise multi-atlas method called Hough-CNN, which was employed to perform the detection and segmentation of multiple deep brain regions. This method used a Hough voting strategy similar to the one proposed in an earlier study [26]; the difference was that the anatomy-specific features were obtained through a CNN instead of through SAEs. To make full use of the contextual information in 3D US volumes, Pourtaherian et al. [148] directly trained a 3D CNN to detect needle voxels in 3D US volumes; each voxel was categorized from locally extracted raw data of three orthogonal planes centered on it. To address the issue of highly imbalanced datasets, a new update strategy involving the informed re-sampling of non-needle voxels in the training stage was adopted in order to improve the detection performance and robustness.

A typical non-rigid object segmentation scheme that is widely applied to 2D images is also suitable for the segmentation of 3D

US volumes. Ghesu et al. [52] employed this typical non-rigid segmentation method, which consists of two steps—rigid object localization and non-rigid object boundary estimation—to achieve the detection and segmentation of the aortic valve in 3D US volumes. To address the issue of 3D object detection, marginal space deep learning (MSDL), which takes advantage of marginal space learning (MSL) [149] and deep learning, was adopted. Based on the detected object, an initial estimation of the non-rigid shape was determined, followed by a sparse adaptive DNN-based active shape model to guild the shape deformation. The results on a large 3D transesophageal echocardiogram image dataset demonstrated the efficiency and robustness of the MSDL in the 3D detection and segmentation task of the aortic valve; it showed a significant improvement of up to 42.5% over the state of the art. By only using the central processing unit (CPU), the aortic valve can be successfully segmented in less than one second with higher accuracy than the original MSL.

The segmentation of fetal structures is more challenging than that of anatomical structures or organs. For example, the placenta is highly variable, as its position depends on the implantation site in the uterus. Although handcrafted manual segmentation and semi-automated methods have proven to be accurate and acceptable, they are time consuming and operator dependent. To address these issues, Looney et al. [150] employed DeepMedic to segment the placenta from 3D US volumes. No manual annotations were used in the training dataset; instead, the output of the semi-automated random walker (RW) method was used as the ground truth. DeepMedic is a dual pathway 3D CNN architecture, proposed by Kamnitsas et al. [151], which was originally used to segment lesions in brain MRI. However, the successful placental segmentation from the 3D US volumes seemed to demonstrate that DeepMedic is a generic 3D deep architecture that is suitable for different modalities of 3D medical data (volumes). Recently, Yang et al. [152] implanted an RNN into the customized 3D FCN for the simultaneous segmentation of multiple objects in US volumes, including fetus, gestational sac, and placenta. To tackle the ubiquitous boundary uncertainty, an effective serialization strategy was adopted. In addition, a hierarchical deep supervision mechanism was proposed to boost the information flow within the RNN and further improve the segmentation performance. Similarly, Schmidt-Richberg et al. [153] integrated the FCN into deformable shape models for 3D fetal abdominal US volume segmentation.

4. Future challenges and perspectives

From the examples provided above, it is evident that deep learning has entered various application areas of medical US analysis. However, although deep learning methods have constantly updated state-of-the-art performance results across different application aspects in medical US analysis, there is still room for improvement. In this section, we summarize the overall challenges commonly encountered in the application of deep learning in medical US analysis, and discuss future perspectives.

Clearly, the major performance improvement that can be achieved with deep learning greatly depends on large training sample datasets. However, compared with the large and publicly available datasets in other areas (e.g., more than 1 million annotated multi-label natural images in ImageNet [6]), the current public availability of datasets in the field of medical US is still limited. The limited training data act as a bottleneck for the further application of deep learning methods in medical US image analysis.

To address the issue of small sample datasets, one of the most commonly used methods by researchers at present is performing cross-dataset (intra-modality or inter-modality) learning—that is, transfer learning. As pointed out earlier, there are two main ideas

regarding the use of transfer learning: directly utilizing a pre-trained network as a feature extractor, and fine-tuning by fixing the weights in parts of the network [77]. Depending on whether the destination and source come from the same domain or not, transfer learning can be divided into two types: cross-modal and cross-domain transfer learning. Cross-domain transfer learning is the most common way to accomplish a variety of tasks in medical US analysis. In any case, the pre-training of models is currently always performed on large sample datasets. Doing so ensures an excellent performance; however, this is absolutely not the optimal choice in the medical imaging domain. When using small training samples, the *de novo* training of domain-specific deep models (if the size of the model is selected properly) can achieve a superior performance when compared with transfer learning from a network that has been pre-trained using large training samples in another domain (e.g., natural images) [154]. The underlying reason for this may be that the mapping from the raw input image pixels to the feature vectors used for a specific task (e.g., classification) in medical imaging is much more complex in the pre-trained case, and requires a large training sample for good generalization. Instead, a specially designed small network may be more suitable to the smaller-scale training datasets that are commonly encountered in medical imaging [155]. Consequently, developing domain-specific deep learning models for medical imaging can not only improve task-specific performance with a low computation complexity, but also facilitate technological advantages in CADx in the medical imaging domain.

In addition, models trained on natural images may not be optimal for medical images, which are typically single channel, low contrast, and texture rich. In medical imaging, and especially in breast imaging, multiple modalities such as MRI, X-ray, and US are frequently used in the diagnostic workflow. Either US or mammography (i.e., X-ray) is usually regarded as the first-line screening examination, for which it is much easier to collect large training samples. However, breast MRI is a more costly and time-consuming method that is commonly used for screening high-risk populations, and it is much more difficult to collect sufficient training datasets and ground-truth annotation for this method. In this case, cross-modal transfer learning can be an advisable choice. Few experiments have demonstrated that cross-modal transfer learning may be superior to a cross-domain one for a specific task in the absence of sufficient training datasets [156]. Considering the fact that large samples are rarely collected from a single site (i.e., institute or hospital), and are instead often collected from multiple different sites (or machines), it is possible to make attempts to perform cross-site (or cross-machine) transfer learning of the same modality.

Finally, other issues regarding current transfer learning algorithms must be addressed; these include how to avoid negative transfer, how to deal with heterogeneous feature spaces between source and target domains or tasks, and how to improve generalization across different tasks. The purpose of transfer learning is to leverage the knowledge learned from the source task in order to improve learning performance in the target task. However, an inappropriate transfer learning method may sometimes decrease the performance instead, resulting in negative transfer [157].

Ignoring the inherent differences between different methods, the effectiveness of any transfer method for a given target task mainly depends on two aspects: the source task, and how it is related to the target. Ideally, a transfer method would produce a positive transfer between sufficiently related tasks while avoiding negative transfer, although the tasks would not be an appropriate match. However, these goals are difficult to achieve simultaneously in practice. To avoid negative transfer, the following strategies may be used: ① recognizing and rejecting harmful source task knowledge, ② choosing the best source task from a set of

candidate source tasks (if possible), and ③ modeling the task similarity between multiple candidate source tasks. In addition, mapping is necessary in order to translate between task representations when the representations of the source and target tasks are heterogeneous.

It is worth stressing again that 3D US is a very important imaging modality in the field of medical imaging, and that 3D US image analysis has shown great potential in US-based clinical application, although several issues remain to be addressed. It can be foreseen that more novel 3D deep learning algorithms will be developed to perform various tasks in medical US analysis, and that greater performance improvements will be achieved in the future. However, it is currently difficult to proceed with the development of 3D deep learning methods without the strong support of other communities, and especially that of the CV community.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61571304, 81571758, and 61701312), in part by the National Key Research and Development Program of China (2016YFC0104703), in part by the Medical Scientific Research Foundation of Guangdong Province, China (B2018031), and in part by the Shenzhen Peacock Plan (KQTD2016053112051497).

Compliance with ethics guidelines

Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2018.11.020>.

References

- [1] Reddy UM, Filly RA, Copel JA. Prenatal imaging: ultrasonography and magnetic resonance imaging. *Obstet Gynecol* 2008;112(1):145–57.
- [2] Noble JA, Boukerrouji D. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging* 2006;25(8):987–1010.
- [3] Salomon LJ, Winer N, Bernard JP, Ville Y. A score-based method for quality control of fetal images at routine second-trimester ultrasound examination. *Prenat Diagn* 2008;28(9):822–7.
- [4] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6):84–90.
- [5] Wang G. A perspective on deep imaging. *IEEE Access* 2016;4:8914–24.
- [6] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;115(3):211–52.
- [7] Salakhutdinov R. Learning deep generative models. *Annu Rev Stat Appl* 2015;2(1):361–85.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [9] Deng L, Yu D. Deep learning: methods and applications. *Found Trends Signal Process* 2014;7(3–4):197–387.
- [10] Deng L, Li J, Huang JT, Yao K, Yu D, Seide F, et al. Recent advances in deep learning for speech research at Microsoft. In: *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2013 May 26–31; Vancouver, BC, Canada. New York: IEEE; 2013. p. 8604–8.
- [11] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19(1):221–48.
- [12] Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 2016;35(5):1153–9.
- [13] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
- [14] Suzuki K. Overview of deep learning in medical imaging. *Radiological Phys Technol* 2017;10(3):257–73.
- [15] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access* 2018;6:9375–89.
- [16] Anas EMA, Seitel A, Rasoulia A, John PS, Pichora D, Darras K, et al. Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. *Int J CARS* 2015;10(6):959–69.
- [17] Hiramatsu Y, Muramatsu C, Kobayashi H, Hara T, Fujita H. Automated detection of masses on whole breast volume ultrasound scanner: false positive reduction using deep convolutional neural network. In: *Proceedings of the SPIE Medical Imaging*; 2017 Feb 11–16; Orlando, FL, USA. Bellingham: SPIE; 2017.
- [18] Bian C, Lee R, Chou Y, Cheng J. Boundary regularized convolutional neural network for layer parsing of breast anatomy in automated whole breast ultrasound. In: *Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, editors. Medical image computing and computer-assisted intervention—MICCAI 2017*. Berlin: Springer; 2017. p. 259–66.
- [19] Shi J, Zhou S, Liu X, Zhang Q, Lu M, Wang T. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. *Neurocomputing* 2016;194:87–94.
- [20] Azizi S, Imani F, Zhuang B, Tahmasebi A, Kwak JT, Xu S, et al. Ultrasound-based detection of prostate cancer using automatic feature selection with deep belief networks. In: *Navab N, Hornegger J, Wells W, Frangi A, editors. Medical image computing and computer-assisted intervention—MICCAI 2015*. Berlin: Springer; 2015. p. 70–7.
- [21] Yang X, Yu L, Wu L, Wang Y, Ni D, Qin J, et al. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*; 2017 Feb 4–9; San Francisco, CA, USA: AAAI Press; 2017. p. 1633–9.
- [22] Wu K, Chen X, Ding M. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik* 2014;125(15):4057–63.
- [23] Ghesu FC, Georgescu B, Zheng Y, Hornegger J, Comaniciu D. Marginal space deep learning: efficient architecture for detection in volumetric image data. In: *Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical image computing and computer-assisted intervention*. Berlin: Springer; 2015. p. 710–8.
- [24] Pereira F, Bueno A, Rodriguez A, Perrin D, Marx G, Cardinale M, et al. Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms. *J Med Imaging* 2017;4(1):014502.
- [25] Sombune P, Phienphanich P, Phuechpanpaisal S, Muengtaeweepong S, Ruamthanthong A, Tantibundhit C. Automated embolic signal detection using deep convolutional neural network. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway: IEEE; 2017. p. 3365–8.
- [26] Milletari F, Ahmadi SA, Kroll C, Hennemperger C, Tombari C, Shah A, et al. Robust segmentation of various anatomies in 3D ultrasound using hough forests and learned data representations. In: *Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical image computing and computer-assisted intervention*. Berlin: Springer; 2015. p. 111–8.
- [27] Lekadir K, Galimzianova A, Betriu A, Del Mar Vila M, Igual L, Rubin DL, et al. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound. *IEEE J Biomed Health Inform* 2017;21(1):48–55.
- [28] Shin J, Tajbakhsh N, Hurst RT, Kendall CB, Liang J. Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 2526–35.
- [29] Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Med Phys* 2017;44(5):1678–91.
- [30] Smistad E, Løvstakken L. Vessel detection in ultrasound images using deep convolutional neural networks. In: *Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep learning and data labeling for medical applications*. Berlin: Springer; 2016. p. 30–8.
- [31] Su S, Gao Z, Zhang H, Lin Q, Hao WK, Li S. Detection of lumen and media-adventitia borders in IVUS images using sparse auto-encoder neural network. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21, Melbourne, Australia. Piscataway: IEEE; 2017. p. 1120–4.
- [32] Yaqub M, Kelly B, Papageorgiou AT, Noble JA. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. *Ultrasound Med Biol* 2017;43(12):2925–33.
- [33] Huang W, Bridge CP, Noble JA, Zisserman A. Temporal HeartNet: towards human-level automatic analysis of fetal cardiac screening video. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Berlin: Springer; 2017. p. 341–9.
- [34] Gao Y, Noble JA. Detection and characterization of the fetal heartbeat in free-hand ultrasound sweeps with weakly-supervised two-streams convolutional networks. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Berlin: Springer; 2017. p. 305–13.
- [35] Qi H, Collins S, Noble A. Weakly supervised learning of placental ultrasound images with residual networks. In: *Hernández MV, González-Castro V, editors. Medical image understanding and analysis*. Berlin: Springer; 2017. p. 98–108.
- [36] Chen H, Zheng Y, Park JH, Heng PA, Zhou K. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In: *Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2016 Oct 17–21; Athens, Greece. Berlin: Springer; 2016. p. 487–95.

- [37] Ravishanker H, Prabhu SM, Vaidya V, Singhal N. Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning. In: *Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging*; 2016 Jun 13–16; Prague, Czech Republic. Piscataway: IEEE; 2016. p. 779–82.
- [38] Zhang Y, Ying MTC, Yang L, Ahuja AT, Chen DZ, et al. Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: *Proceedings of 2016 IEEE International Conference on Bioinformatics and Biomedicine*; 2016 Dec 15–18; Shenzhen, China. Piscataway: IEEE; 2016. p. 443–8.
- [39] Ravishanker H, Venkataramani R, Thiruvankadam S, Sudhakar P, Vaidya V. Learning and incorporating shape models for semantic segmentation. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 203–11.
- [40] Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R. SLIDE: automatic spine level identification system using a deep convolutional neural network. *Int J CARS* 2017;12(7):1189–98.
- [41] Golan D, Donner Y, Mansi C, Jaremko J, Ramachandran M. Fully automating Graf's method for DDH diagnosis using deep convolutional neural networks. In: *Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep learning and data labeling for medical applications. Proceedings of International Workshops on DLMIA and LABELS*; 2016 Oct 21; Athens, Greece. Berlin: Springer; 2016. p. 130–41.
- [42] Hareendranathan AR, Zonoobi D, Mabee M, Cobzas D, Punithakumar K, Noga ML, et al. Toward automatic diagnosis of hip dysplasia from 2D ultrasound. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21; Melbourne, Australia. 2017. p. 982–5.
- [43] Burlina P, Billings S, Joshi N, Albayda J. Automated diagnosis of myositis from muscle ultrasound: exploring the use of machine learning and deep learning methods. *PLoS ONE* 2017;12(8):e0184059.
- [44] Hafiane A, Vieyres P, Delbos A. Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images. 2017. arXiv:1706.05870.
- [45] Fasel I, Berry J. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In: *Proceedings of 2010 20th International Conference on Pattern Recognition*; 2010 Aug 23–26; Istanbul, Turkey. p. 1493–6.
- [46] Jaumard-Hakoun A, Xu K, Roussel-Ragot P, Dreyfus G, Denby B. Tongue contour extraction from ultrasound images based on deep neural network. 2016. arXiv:1605.05912.
- [47] Xu K, Roussel P, Csapó TG, Denby B. Convolutional neural network-based automatic classification of midsagittal tongue gestural targets using B-mode ultrasound images. *J Acoust Soc Am* 2017;141(6):EL531–7.
- [48] Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging* 2017;30(4):477–86.
- [49] Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J Digit Imaging* 2017;30(2):234–43.
- [50] Li Y, Xu R, Ohya J, Iwata H. Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers. In: *Proceedings of 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 2017 Jul 11–15; Seogwipo, Korea. Piscataway: IEEE; 2017. p. 1485–8.
- [51] Fenster A, Downey DB, Cardinal HN. Three-dimensional ultrasound imaging. *Phys Med Biol* 2001;46(5):R67–99.
- [52] Ghesu FC, Krubasik E, Georgescu B, Singh V, Zheng Y, Hornegger J, et al. Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE Trans Med Imaging* 2016;35(5):1217–28.
- [53] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 2017;30(4):449–59.
- [54] Xing F, Xie Y, Su H, Liu F, Yang L. Deep learning in microscopy image analysis: a survey. *IEEE Trans Neural Networks Learn Syst* 2017;29(10):4550–68.
- [55] Xian M, Zhang Y, Cheng HD, Xu F, Zhang B, Ding J. Automatic breast ultrasound image segmentation: a survey. *Pattern Recognit* 2018;79:340–55.
- [56] He K, Sun J. Convolutional neural networks at constrained time cost. In: *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*; 2015 Oct 7–12; Boston, MA, USA. Piscataway: IEEE; 2015. p. 5353–60.
- [57] Milletari F, Ahmadi SA, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput Vis Image Underst* 2017;164:92–102.
- [58] Liu X, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors* 2017;17(1):149.
- [59] Canziani A, Paszke A, Culurciello E. An analysis of deep neural network models for practical applications. 2016. arXiv:1605.07678.
- [60] Chen H, Dou Q, Ni D, Cheng J, Qin J, Li S, et al. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: *Proceedings of 2015 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2015 Oct 5–9; Munich, Germany. Berlin: Springer; 2015. p. 507–14.
- [61] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 1994;5(2):157–66.
- [62] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
- [63] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014. arXiv:1406.1078.
- [64] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013;35(8):1798–828.
- [65] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 2010;11(12):3371–408.
- [66] Hinton GE. A practical guide to training restricted boltzmann machines. In: *Montavon G, Orr GB, Müller KR, editors. Neural networks: tricks of the trade*. Berlin: Springer; 2012. p. 599–619.
- [67] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313(5786):504–7.
- [68] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 2016;35(5):1299–312.
- [69] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 2011;12(7):2121–59.
- [70] Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: *Proceedings of the 30th International Conference on International Conference on Machine Learning*; 2013 Jun 16–21; Atlanta, GA, USA. JMLR; 2013. p. 1139–47.
- [71] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*; 2010 Jun 21–24; Haifa, Israel. Piscataway: Omnipress; 2010. p. 807–14.
- [72] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*; 2011 Apr 11–13; Ft. Lauderdale, FL, USA. JMLR; 2011. p. 315–23.
- [73] Goodfellow IJ, Warde-Farley D, Mirza M, Courville A, Bengio Y. Maxout networks. In: *Proceedings of the 30th International Conference on Machine Learning*; 2013 Jun 16–21; Atlanta, GA, USA. JMLR; 2013. p. 1319–27.
- [74] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. 2012. arXiv: 1207.0580.
- [75] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning*; 2015 Jul 6–11; Lille, France. JMLR; 2015. p. 448–56.
- [76] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345–59.
- [77] Azizi S, Mousavi P, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection. *Int J CARS* 2017;12(7):1111–21.
- [78] Chen H, Ni D, Qin J, Li S, Yang X, Wang T, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Health Inform* 2015;19(5):1627–36.
- [79] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*; 2014 Nov 3–7; New York, NY, USA. New York: ACM; 2014. p. 675–8.
- [80] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. 2016. arXiv: 1603.04467.
- [81] Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, et al. Theano: new features and speed improvements. 2012. arXiv: 1211.5590.
- [82] Collobert R, Kavukcuoglu K, Farabet C. Torch7: a matlab-like environment for machine learning. In: *Proceedings of the NIPS 2011 Workshop*; 2011 Dec 12–17; Granada, Spain, 2011.
- [83] Chen T, Li M, Li Y, Lin M, Wang N, Wang M. MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. 2015. arXiv: 1512.01274.
- [84] Bahrampour S, Ramakrishnan N, Schott L, Shah M. Comparative study of deep learning software frameworks. 2015. arXiv: 1511.06435.
- [85] Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med Phys* 2008;35(12):5799–820.
- [86] Jamieson A, Drukker K, Giger M. Breast image feature learning with adaptive deconvolutional networks. In: *Proceedings of the SPIE Medical Imaging*; 2012 Feb 4–9; San Diego, CA, USA. Bellingham: SPIE; 2012.
- [87] Liu X, Shi J, Zhang Q. Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset. In: *Proceedings of the 6th International Workshop on Machine Learning in Medical Imaging*; 2015 Oct 5; Munich, Germany. Berlin: Springer; 2015. p. 313–20.
- [88] Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep* 2016;6(1):24454.
- [89] Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* 2016;72:150–7.
- [90] Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol* 2017;62(19):7714–28.
- [91] Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys* 2017;44(10):5162–71.

- [92] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127(12):2893–917.
- [93] Guo L, Wang D, Xu H, Qian Y, Wang C, Zheng X, et al. CEUS-based classification of liver tumors with deep canonical correlation analysis and multi-kernel learning. In: *Proceedings of 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 2017 Jul 11–15; Seogwipo, Korea. Piscataway: IEEE; 2017. p. 1748–51.
- [94] Meng D, Zhang L, Cao G, Cao W, Zhang G, Hu B. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access* 2017;5:5804–10.
- [95] Ma J, Wu F, Zhu J, Xu D, Kong D. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* 2017;73:221–30.
- [96] Liu T, Xie S, Yu J, Niu L, Sun WD. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. In: *Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2017 Jun 19; New Orleans, LA, USA. Piscataway: IEEE; 2017. p. 919–23.
- [97] Liu T, Xie S, Zhang Y, Yu J, Niu L, Sun W. Feature selection and thyroid nodule classification using transfer learning. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 1096–9.
- [98] Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 2002;19(2):190–6.
- [99] Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. *IEEE Trans Cybern* 2017;47(5):1336–49.
- [100] Jang J, Kwon JY, Kim B, Lee SM, Park Y, Seo JK. CNN-based estimation of abdominal circumference from ultrasound images. 2017. arXiv: 1702.02741.
- [101] Gao Y, Maraci MA, Noble JA. Describing ultrasound video content using deep convolutional neural networks. In: *Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging*. Piscataway: IEEE; 2016. p. 787–90.
- [102] Sundaresan V, Bridge CP, Ioannou C, Noble A. Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: *2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 671–4.
- [103] Perrin DP, Bueno A, Rodriguez A, Marx GR, Del Nido PJ. Application of convolutional artificial neural networks to echocardiograms for differentiating congenital heart diseases in a pediatric population. In: *Proceedings of the SPIE Medical Imaging 2017: Computer-aided Diagnosis*; 2017 Mar 3; Orlando, FL, USA. Bellingham: SPIE; 2012.
- [104] Yu Z, Ni D, Chen S, Li S, Wang T, Lei B. Fetal facial standard plane recognition via very deep convolutional networks. In: *Proceedings of 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 2016 Aug 16–20; Orlando, FL, USA. Piscataway: IEEE; 2016. p. 627–30.
- [105] Yu Z, Tan EL, Ni D, Qin J, Chen S, Li S, et al. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. *IEEE J Biomed Health Inform* 2018;22(3):874–85.
- [106] Azizi S, Imani F, Ghavidel S, Tahmasebi A, Kwak JT, Xu S, et al. Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. *Int J CARS* 2016;11(6):947–56.
- [107] Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. *Int J CARS* 2017;12(8):1293–305.
- [108] Yap MH, Pons G, Martí J, Ganau S, Sentis M, Zwigelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* 2018;22(4):1218–26.
- [109] Cao Z, Duan L, Yang G, Yue T, Chen Q, Fu C, et al. Breast tumor detection in ultrasound images using deep learning. In: *Wu G, Munsell B, Zhan Y, Bai W, Sanroma G, Coupé P, editors. Patch-based techniques in medical imaging*. Berlin: Springer; 2017. p. 121–8.
- [110] Girshick R. Fast R-CNN. In: *Proceedings of 2015 IEEE International Conference on Computer Vision*; 2015 Dec 7–13; Santiago, Chile. Piscataway: IEEE; 2015. p. 1440–8.
- [111] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39(6):1137–49.
- [112] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 779–88.
- [113] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, et al. SSD: single shot multibox detector. In: *Proceedings of the European Conference on Computer Vision*; 2016 Oct 11–14; Amsterdam, The Netherlands. Berlin: Springer; 2016. p. 21–37.
- [114] Lipsanen A, Parkkinen S, Khabbal J, Mäkinen P, Peräniemi S, Hiltunen M, et al. KB-R7943, an inhibitor of the reverse Na⁺/Ca²⁺ exchanger, does not modify secondary pathology in the thalamus following focal cerebral stroke in rats. *Neurosci Lett* 2014;580:173–7.
- [115] Yang X, Ni D, Qin J, Li S, Wang T, Chen S, et al. Standard plane localization in ultrasound by radial component. In: *Proceedings of 2014 IEEE 11th International Symposium on Biomedical Imaging*; 2014 Apr 29–May 2; Beijing, China. Piscataway: IEEE; 2014. p. 1180–3.
- [116] Ni D, Li T, Yang X, Qin J, Li S, Chin C, et al. Selective search and sequential detection for standard plane localization in ultrasound. In: *Yoshida H, Warfield S, Vannier MW, editors. Abdominal imaging, computation and clinical applications*. Berlin: Springer; 2013. p. 203–11.
- [117] Baumgartner CF, Kamnitsas K, Matthew J, Smith S, Kainz B, Rueckert D, et al. Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: *Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2016 Oct 17–21; Athens, Greece. Berlin: Springer; 2016. p. 203–11.
- [118] Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans Med Imaging* 2017;36(11):2204–15.
- [119] Chen H, Ni D, Yang X, Li S, Heng PA. Fetal abdominal standard plane localization through representation learning with knowledge transfer. In: *Wu G, Zhang D, Zhou L, editors. Machine learning in medical imaging*. Berlin: Springer; 2014. p. 125–32.
- [120] Chen H, Wu L, Dou Q, Qin J, Li S, Cheng JZ, et al. Ultrasound standard plane detection using a composite neural network framework. *IEEE Trans Cybern* 2017;47(6):1576–86.
- [121] Dezaki FT, Dhungel N, Abdi A, Luong C, Tsang T, Jue J, et al. Deep residual recurrent neural networks for characterization of cardiac cycle phase from echocardiograms. In: *Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, editors. Deep learning in medical image analysis and multimodal learning for clinical decision support*. Berlin: Springer; 2017. p. 100–8.
- [122] Sofka M, Milletari F, Jia J, Rothberg A. Fully convolutional regression network for accurate detection of measurement points. In: *Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, editors. Deep learning in medical image analysis and multimodal learning for clinical decision support*. Berlin: Springer; 2017. p. 258–66.
- [123] Ghesu FC, Georgescu B, Mansi T, Neumann D, Hornegger J, Comaniciu D. An artificial agent for anatomical landmark detection in medical images. In: *Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2016 Oct 17–21; Athens, Greece. Berlin: Springer; 2016. p. 229–37.
- [124] Nascimento JC, Carneiro G. Multi-atlas segmentation using manifold learning with deep belief networks. In: *Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging*; 2016 Apr 13–16; Prague, Czech Republic. Piscataway: IEEE; 2016. p. 867–71.
- [125] Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int J CARS* 2017;12(11):1895–910.
- [126] Singhal N, Mukherjee S, Perrey C. Automated assessment of endometrium from transvaginal ultrasound using Deep Learned Snake. In: *Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 83–6.
- [127] Bernard O, Touil B, Gelas A, Prost R, Friboulet D. A RBF-Based multiphase level set method for segmentation in echocardiography using the statistics of the radiofrequency signal. In: *Proceedings of 2007 IEEE International Conference on Image Processing*; 2007 Oct 16–19; San Antonio, TX, USA. Piscataway: IEEE; 2007.
- [128] Jacob G, Noble JA, Behrenbruch C, Kelion AD, Banning AP. A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. *IEEE Trans Med Imaging* 2002;21(3):226–38.
- [129] Carneiro G, Nascimento J, Freitas A. Robust left ventricle segmentation from ultrasound data using deep neural networks and efficient search methods. In: *Proceedings of 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*; 2010 Apr 14–17; Rotterdam, The Netherlands. Piscataway: IEEE; 2010. p. 1085–8.
- [130] Carneiro G, Nascimento JC. Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures. In: *Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2010 Jun 13–18. San Francisco, CA, USA. Piscataway: IEEE; 2010. p. 2815–22.
- [131] Carneiro G, Nascimento JC. Incremental on-line semi-supervised learning for segmenting the left ventricle of the heart from ultrasound data. In: *Proceedings of 2011 International Conference on Computer Vision*; 2011 Nov 6–13; Barcelona, Spain. Piscataway: IEEE; 2011. p. 1700–7.
- [132] Carneiro G, Nascimento JC. The use of on-line co-training to reduce the training set size in pattern recognition methods: application to left ventricle segmentation in ultrasound. In: *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*; 2012 Jun 16–21; Providence, RI, USA. Piscataway: IEEE; 2012. p. 948–55.
- [133] Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Trans Image Process* 2012; 21(3):968–82.
- [134] Carneiro G, Nascimento JC. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Trans Pattern Anal Mach Intell* 2013;35(11):2592–607.
- [135] Nascimento JC, Carneiro G. Deep learning on sparse manifolds for faster object segmentation. *IEEE Trans Image Process* 2017;26(10):4978–90.
- [136] Nascimento JC, Carneiro G. Non-rigid segmentation using sparse low dimensional manifolds and deep belief networks. In: *Proceedings of 2014*

- IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. Piscataway: IEEE; 2010. p. 288–95.
- [137] Raynaud C, Langet H, Amzulescu MS, Saloux E, Bertrand H, Allain P, et al. Handcrafted features vs. ConvNets in 2D echocardiographic images. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 1116–9.
- [138] Xiong X, Torre FDL. Global supervised descent method. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition; 2015 Jun 7–12; Boston, MA, USA. Piscataway: IEEE; 2015. p. 2664–73.
- [139] Yu L, Guo Y, Wang Y, Yu J, Chen P. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. *IEEE Trans Biomed Eng* 2017;64(8):1886–95.
- [140] Baka N, Leenstra S, van Walsum T. Ultrasound aided vertebral level localization for lumbar surgery. *IEEE Trans Med Imaging* 2017;36(10):2138–47.
- [141] Wu L, Xin Y, Li S, Wang T, Heng P, Ni D. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 663–6.
- [142] Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2010;32(10):1744–57.
- [143] Anas EMA, Nouranian S, Mahdavi SS, Spadinger I, Morris WJ, Salcudean SE, et al. Clinical target-volume delineation in prostate brachytherapy using residual neural networks. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 365–73.
- [144] Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D. 3D deep learning for efficient and robust landmark detection in volumetric data. In: Proceedings of 2015 IEEE Medical Image Computing and Computer-Assisted Intervention; 2015 Oct 5–9; Munich, Germany. Piscataway: IEEE; 2015. p. 565–72.
- [145] Pesteie M, Lessoway V, Abolmaesumi P, Rohling RN. Automatic localization of the needle target for ultrasound-guided epidural injections. *IEEE Trans Med Imaging* 2018;37(1):81–92.
- [146] Nie S, Yu J, Chen P, Wang Y, Zhang JQ. Automatic detection of standard sagittal plane in the first trimester of pregnancy using 3-D ultrasound data. *Ultrasound Med Biol* 2017;43(1):286–300.
- [147] Nie S, Yu J, Chen P, Zhang J, Wang Y. A novel method with a deep network and directional edges for automatic detection of a fetal head. In: Proceedings of 2015 the 23rd European Signal Processing Conference; 2015 Aug 31–Sep 4; Nice, France. Piscataway: IEEE; 2015. p. 654–8.
- [148] Pourtaherian A, Zanjani FG, Zinger S, Mihajlovic N, Ng G, Korsten H, et al. Improving needle detection in 3D ultrasound using orthogonal-plane convolutional networks. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 610–8.
- [149] Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans Med Imaging* 2008;27(11):1668–81.
- [150] Looney P, Stevenson GN, Nicolaides KH, Plasencia W, Molloholli M, Natsis S, et al. Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 279–82.
- [151] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61–78.
- [152] Yang X, Yu L, Li S, Wang X, Wang N, Qin J, et al. Towards automatic semantic segmentation in volumetric ultrasound. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 711–9.
- [153] Schmidt-Richberg A, Brosch T, Schadevaldt N, Klinder T, Caballero A, Salim I, et al. Abdomen segmentation in 3D fetal ultrasound using CNN-powered deformable models. In: Proceedings of the 4th International Workshop on Fetal and Infant Image Analysis; 2017 Sep 14; Quebec City, Canada. Piscataway: IEEE; 2017. p. 52–61.
- [154] Amit G, Ben-Ari R, Hadad O, Monovich E, Granot N, Hashoul S. Classification of breast MRI lesions using small-size training sets: comparison of deep learning approaches. In: Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis; 2017 Mar 3; Orlando, Florida. Bellingham: SPIE; 2017.
- [155] Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35(5):1285–98.
- [156] Hadad O, Bakalo R, Ben-Ari R, Hashoul S, Amit G. Classification of breast lesions using cross-modal deep learning. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 109–12.
- [157] Lisa T, Jude S. Transfer learning. In: Olivas ES, Guerrero JDM, Martinez-Sober M, Magdalena-Benedito JR, López AJ, editors. Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. Hershey: IGI Global; 2010. p. 242–64.



Research
AI for Precision Medicine—Review

深度学习在医学超声图像分析中的应用综述

刘盛锋^a, 王毅^a, 杨鑫^b, 雷柏英^a, 刘立^a, 李享^a, 倪东^{a,*}, 汪天富^{a,*}

^a National-Regional Key Technology Engineering Laboratory for Medical Ultrasound & Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging & School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China

^b Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:

Received 1 March 2018

Revised 18 July 2018

Accepted 8 November 2018

Available online 29 January 2019

关键词

深度学习

医学超声图像分析

分类

分割

检测

摘要

在临床上, 超声(US)已成为最主要的成像模态之一。超声是一种快速发展的技术, 具有无痛苦、无电离辐射、经济适用、实时成像等优点; 同时也有成像质量差、差异性大等特有的缺点。对于图像分析来说, 很有必要开发先进的自动化超声图像分析方法来帮助医生进行超声诊断, 一方面可以减轻医生的负担, 另一方面可以降低诊断的主观性, 从而使得诊断更加客观与准确。近年来, 深度学习已经成为最主要的机器学习工具, 并且广泛应用于各个研究领域, 尤其是一般的图像分析与计算机视觉。在医学超声图像分析中, 深度学习也展示了巨大的应用潜力。本文首先简要介绍了一些流行的深度学习结构, 然后总结并全面讨论了深度学习在超声图像分析的各种特定任务(如图像分类、物体检测与目标分割)中的应用。最后, 本文讨论了深度学习在医学超声图像分析应用中所面临的挑战以及潜在的发展趋势。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

作为最常用的成像模态之一, 超声(US)已经成为临床上不可或缺的扫查与诊断工具。由于具有相对安全、成本较低、无创、实时成像以及操作舒适等优点, 超声成像尤其广泛应用于世界各地的产前筛查中[1]。几十年的临床实践证实, 与其他如X射线、磁共振成像(MRI)与计算机断层扫描(CT)在内的医学成像模态相比, 超声成像具有一些明显优势, 如无电离辐射、携带方便、易于接近以及成本效益高等。在当前的临床实践中, 医学超声已在各个专业科室得到了应用, 如心电图、乳腺超声、腹部超声、经直肠超声、心血管超声以

及产前诊断超声, 尤其广泛应用于妇产科(OB-GYN)[2]。但是, 超声也存在一些独特的挑战, 如因噪声与伪影引起的低成像质量, 高度依赖操作者或者诊断医生的经验, 以及不同机构和厂家超声系统之间具有观测者内或者不同观测者之间的高度差异性。例如, 一项使用超声图像进行产前异常检测的研究表明, 不同医学机构之间的敏感度范围为27.5%~96%[3]。为了应对这些挑战, 开发先进的自动化超声图像分析方法, 能够使得超声诊断和(或)评价, 以及图像引导的干预和(或)治疗更加客观、准确与智能。

深度学习是机器学习的一个分支, 也是一种表征学习方法, 能够直接处理原始数据(如超声图像), 并从

* Corresponding author.

E-mail address: nidong@szu.edu.cn (D. Ni), tfwang@szu.edu.cn (T. Wang).

2095-8099/© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

英文原文: Engineering 2019, 5(2): 261–275

引用本文: Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, Tianfu Wang. Deep Learning in Medical Ultrasound Analysis: A Review. Engineering, <https://doi.org/10.1016/j.eng.2018.11.020>

中自动学习到中层与高层抽象特征。它有潜力进行各种超声图像任务的自动分析,如病变/结节分类、组织分割与目标检测等。自从AlexNet [4]——一种深度卷积神经网络(CNN)与深度学习方法的代表,赢得了2012年ImageNet大规模视觉识别挑战赛(ILSVRC)之后,深度学习开始在机器学习领域引起关注。一年后,深度学习入选世界十大突破技术之一[5],这进一步稳固了其作为各种研究领域,尤其在一般图像分析(包括自然图像与医学图像)与计算机视觉(CV)任务中,作为最主要机器学习工具的地位。到目前为止,在网络结构或者模型方面,深度学习获得了快速的发展,如更深的网络结构[6]与深度产生式模型[7]。同时,深度学习也成功应用于各种研究领域,如CV[8]、自然语言处理(NLP)[9]、语音识别[10]以及医学图像分析[11–15],因此表明深度学习能够在各种自动分析任务中获得显著的性能改善,并能取得最优的性能表现。

当前,深度学习已应用在医学图像分析的各种任务中,如包括分类、分割、检测、配准、生物学测量与质量控制/评价在内的传统诊断任务,以及包括图像引导的干预与治疗在内的新兴任务[16](图1)。在这些任务中,分类、检测与分割是3个最基本的任务。它们被广泛应用于不同解剖结构(器官或身体部位)的医学超声

图像分析中,如乳腺[17,18]、前列腺[19–21]、肝脏[22]、心脏[23,24]、大脑[25,26]、颈动脉[27,28]、甲状腺[29]、心血管[30,31]、胎儿[32–37]、淋巴结[38]、肾脏[39]、脊柱[40]、骨骼[41,42]、皮肤[43]、神经结构[44]、舌[45–47]等。多种深度网络应用于这些任务中。CNN是最流行的深度结构之一,在各种任务中获得了极大的成功,如图像分类[48,49]、物体检测[29,30]与目标分割[44,50]。最常用的方法是使用CNN模型从采集的原始数据(如超声图像)中进行自动学习,从而产生分层抽象表示,然后再通过softmax层或者其他线性分类器[如支持向量机(SVM)]产生一个或者多个概率图或类别标签。在这种情况下,图像标注或标签是必不可少的。这就是所谓的“监督学习”。无监督学习也能从原始数据中学习表示[8,9]。自编码器(AE)与限制玻尔兹曼机(RBM)是医学超声图像分析中最普遍使用的无监督神经网络,取得了不错的性能改善。与监督学习相比,无监督学习有一个显著的优势,它不需要耗时耗力、枯燥乏味以及昂贵的人工标注。

尽管当前的医学超声图像分析主要集中在二维(2D)超声图像处理,深度学习在三维(3D)医学超声图像分析中的应用也呈现了增长的趋势。在过去20多年中,商业公司与研究人员一起极大地促进了3D超声成

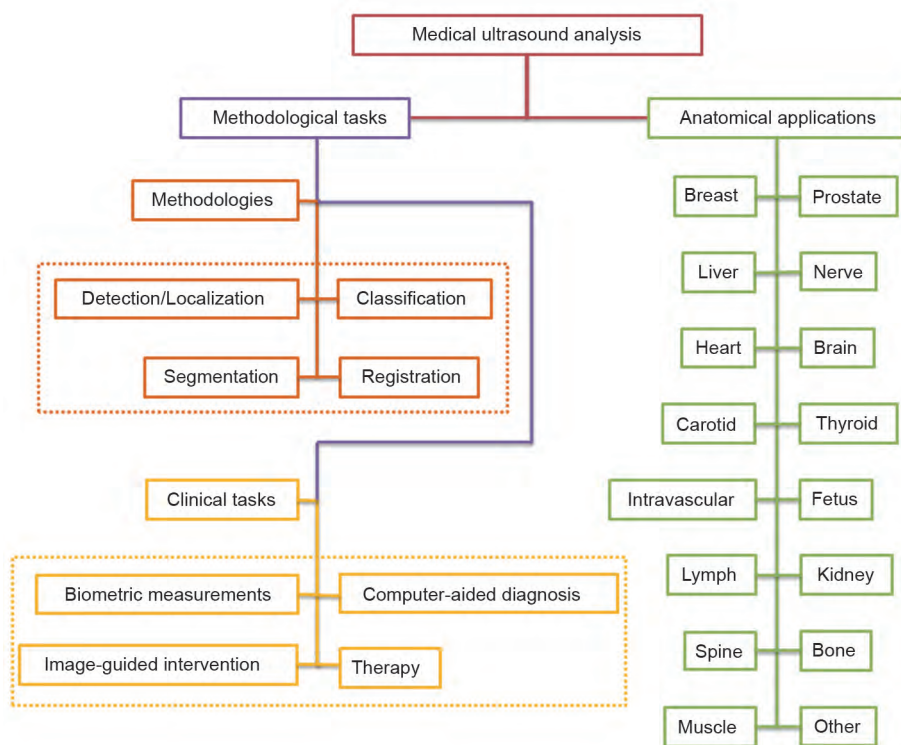


图1. 深度学习在医学超声图像分析中的应用。

像技术的发展。一张3D图像（也常称为“3D体积”）通常比2D图像包含更丰富的信息，因此使用3D体积能够获得更鲁棒的结果。更具体地说，2D超声图像具有一些不可避免的局限性：①尽管超声图像是2D的，而解剖结构是3D的，因此检测人员或者诊断医生在他/她的脑海里一定要有集成多张图像（这个过程通常可能无效且很耗时）的能力，这个能力的缺失将导致较大的差异甚至误诊。②诊断（如妇产科）与治疗（如分期与计划）决策通常需要对器官或肿瘤体积进行准确的估计，然而2D超声通过假定测量目标为一个理想的形状（如椭球体），从长、宽、高等简单的测量中进行体积测量。这可能导致准确率较低，差异性较大，且依赖操作者的经验。③一张2D超声图像呈现了身体中任意角度的一个平面，对于随访研究，这些平面很难再定位与复现[51]。为了克服2D超声的局限性，各种3D超声扫描、重建与显示技术被开发，这给3D超声图像分析提供了很好的基础。此外，3D超声成像技术的发展也支持了当前的深度学习在医学超声图像分析中的应用[23,52]。

到目前为止，已有一些有关深度学习应用于医学图像分析的综述，它们要么涉及整个医学图像分析领域[11–15]，要么只关注单个成像模态，如磁共振成像（MRI）[53]与显微成像[54]。然而，除了极少数涉及特

定的任务，如乳腺超声图像分割[55]以外，很少有文献总结深度学习在医学超声图像分析中的应用。通过指定关键词（如“超声”或“超声检查”或“超声成像”与“卷积”或“深度学习”），在主要的数据库（如PubMed与谷歌学术数据库）和一些重要的会议（如MICCAI、SPIE、ISBI与EMBC）中检索了2018年2月1日之前的这个领域中的所有相关文献。为了筛查所有检索到的文献，通过详细浏览每一篇文献的摘要，保留相关的文献，最后得到近100篇文献，汇总于表2与Appendix A中的Table S1。本文旨在全面而系统地总结深度学习在医学超声图像分析中的应用，主要集中在典型的任务及其在不同解剖结构中的应用。本文的其余部分安排如下：第2节简要介绍医学超声图像分析中常用深度学习方法的理论与结构；第3节详细讨论深度学习在医学超声图像分析中的应用，主要集中在包括分类、检测与分割在内的传统任务；最后，第4节指明深度学习在医学超声图像分析应用中的未来趋势与发展方向。

2. 常用深度学习结构

本节主要介绍医学超声图像分析中普遍使用的深度学习结构。作为机器学习的一个分支，深度学习实质上

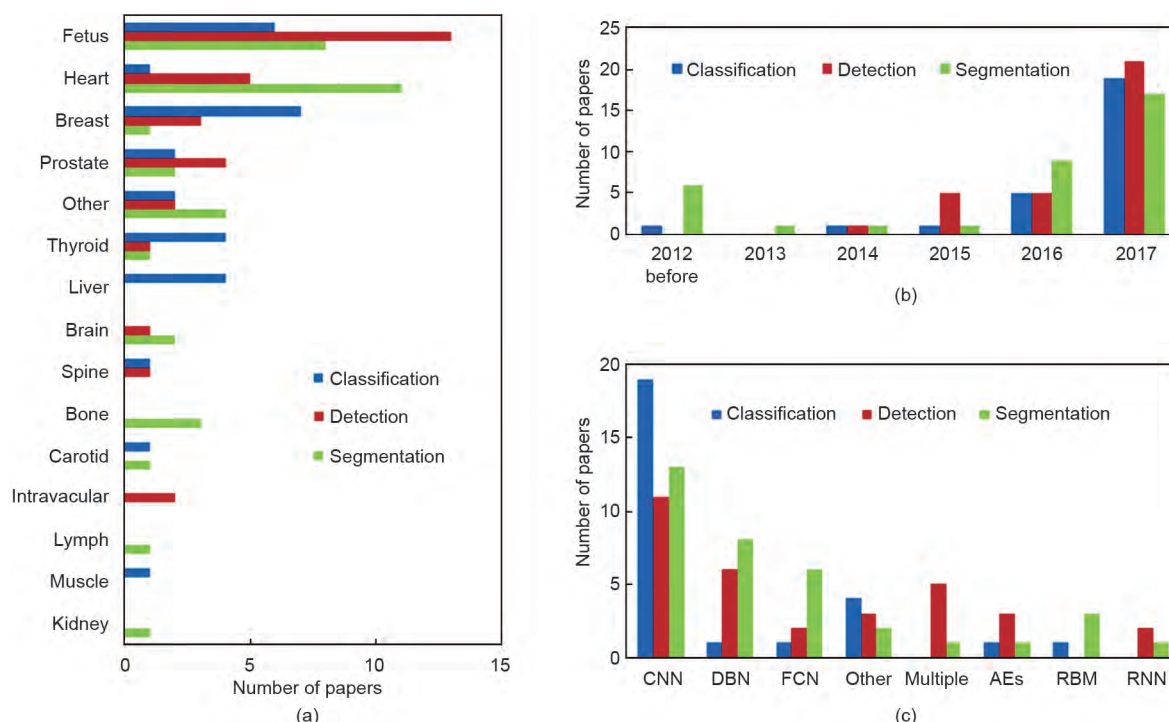


图2. 深度学习在医学超声图像分析中的应用统计。(a) 解剖结构；(b) 出版年份；(c) 网络结构。DBN：深度置信网络；FCN：全卷积网络；Multiple：多种网络结构的组合；RNN：循环神经网络；AE：自编码器[包括它的变体：稀疏自编码器（SAE）和堆叠去噪自编码器]；RBM：限制玻尔兹曼机。

是计算样本数据的分层特征或表示形成, 其中高层次的抽象特征是通过将它们与低级特征相结合来定义的[9]。基于所讨论的深度学习结构与技术(如分类、分割或检测), 本领域最常用的深度学习结构可分为3大类: ①监督深度网络或深度辨别式模型; ②无监督深度网络或深度产生式模型, 以及③混合深度网络。当前医学超声图像分析应用中最基本的模型或结构是CNN、循环神经网络(RNN)、RBM/深度置信网络(DBM)、AE, 以及这些深度学习结构的变种, 如图3所示。上述第三类中的“Multiple”指的是要么由产生式和辨别式模型成分组成, 要么充分利用它们所生成的模型, 因此这一类在下文中不再特别讨论。相反, 本节将继续介绍医学超声图像分析中普遍存在的训练深度模型所遇到的挑战及其应对策略。为了方便起见, 2.4节还总结了一些常用的深度学习框架。

2.1. 有监督深度学习模型

当前, 监督深度学习模型广泛应用于医学图像分析中解剖结构的分类、分割与检测。在这些任务中, CNN与RNN是两个最常用的结构。下面简要介绍一下这两个深度模型。

2.1.1. 卷积神经网络

CNN是一种辨别式深度结构, 包括一些模块, 每一个模块一般由一个卷积层与一个池化层构成。如果有必要的话, 后面还有其他层, 如校正线性单元(ReLU)

与批量标准化(BN)。网络的最后部分一般是全连接层, 这样就构成了一个标准的多层神经网络。就结构而言, 这些模块通常是一块一块堆叠起来, 形成一个深度模型, 这样能够充分利用输入的2D或3D图像的空间与配置信息[8]。

通过对输入图像进行卷积操作, 卷积层共享所有的权重。事实上, 卷积层的作用就是检测输入图像/特征图谱(如医学超声图像)中不同位置的局部特征, 得到一组 k 个核的权重 $W=\{W_1, W_2, \dots, W_k\}$ 以及偏差 $b=\{b_1, b_2, \dots, b_k\}$, 从而产生一个新的特征图谱 A_k^l 。在数学上, 每个卷积层的卷积过程表达为:

$$A_k^l = \sigma(W_k^l * A^{l-1} + b_k^l) \quad (1)$$

式中, $\sigma(\cdot)$ 是非线性激活函数; b_k^l 是偏置参数; $*$ 表示卷积操作。

在一般的CNN模型中, 为了克服卷积过程的减少, 卷积层中超参数的确定非常关键。这主要涉及3个超参数: 深度、步长(stride)与填充(padding)。输出体积的深度对应滤波器的数量, 每一个滤波器学习在输入中寻找局部不同。指定步长能够控制滤波器如何在输入体积中进行卷积操作。在实践中, 更小的步长总是工作得更好, 这是因为在网络的前面一些层(即那些更靠近输入数据的层)中小的步长能够产生大的激活图, 能够产生更好的性能[56]。在一个拥有许多卷积层的CNN中, 由于每一次卷积操作都会丢失一些区域, 尤其是边界, 因此输出维度的减少可能出现。在输入体积的边界

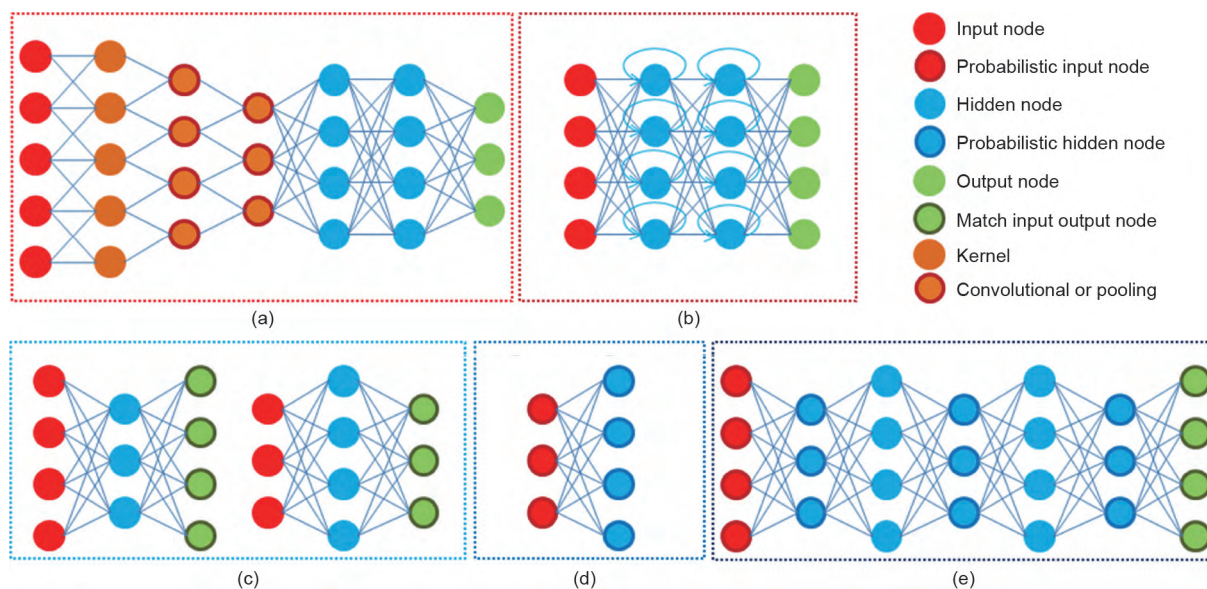


图3. 5种典型的神经网络结构, 主要分为两大类: (1) 监督深度学习模型, 包括 (a) CNN和 (b) RNN; (2) 无监督深度学习模型, 包括 (c) AE/SAE, (d) RBM, 以及 (e) DBM。

周围进行填充（一般是0填充）是消除卷积操作过程中维度降低影响的一种最常用的策略。填充最大的好处之一是它使得设计更深的网络成为了可能。此外，填充防止了输入体积中边界信息的丢失，所以它可以有效改善整个模型的性能。因此，在有限的计算成本与时间成本的条件下，在实践中对特定的任务，有必要在多个因子（即滤波器数量、滤波器大小、步长、网络深度等）之间进行平衡。

卷积层的输出被随后的池化层进行再采样，以降低下面层的数据率。与适当选择的池化方案一起，卷积层中共享的权重能使得CNN具有某些不变性，如平移不变性。这也能大大降低参数的数量，如权重的数量不再完全取决于输入图像的大小。值得注意的是，全连接层一般放在网络中卷积流的后面，通常没有共享权重。在标准的CNN模型中，通过网络最后一层的softmax函数产生激活来获得关于类标签的分布。然而，有时也使用一些传统的机器学习方法，如投票策略[57]或线性SVM[58]。

随着流行度以及实践性增加，许多经典的以及基于CNN的深度学习结构被开发并应用于（医学）图像分析、自然语言处理以及语音识别中。如AlexNet（或适用于Caffe深度学习框架的CaffeNet）、LeNet、faster R-CNN、GoogLeNet、ResNet以及VGGNet。这些网络结构在不同性能指标（如准确率、推理时间、内存以及参数使用）上的详细比较请参考文献[59]。

2.1.2. 循环神经网络

在实践中，RNN一般作为一种监督深度学习网络应用于医学超声图像分析的各种任务中[21,60]。在RNN中，网络的深度可以与输入样本数据序列（如医学超声视频序列）的长度一样长。一个普通的RNN包含一个潜藏或隐藏状态， h_t ，它是时间 t 时由网络输入 x_t 与前一个状态 h_{t-1} 进行非线性映射得到的输出，可以表示为：

$$h_t = \sigma(Wx_t + Rh_{t-1} + b) \quad (2)$$

式中，权重 W 与 R 在时间上是共享的； b 是偏差参数。

由于结构特性，RNN在建模序列数据（如医学超声视频序列）时具有先天的优势。但是，到目前为止，RNN都没有广泛应用于各种涉及序列建模的研究任务中。部分原因是很难训练RNN来捕捉长期依赖性，这种情况下，RNN通常引发梯度爆炸或梯度消失，这个问题早在20世纪90年代就被发现了[61]。因此，一些特别的

记忆单元得到开发，最早的以及最常用的是长短时记忆细胞（LSTM）[62]及其简化门控循环单元[63]。到目前为止，RNN主要应用于语音或文本识别领域，在医学图像分析中应用较少，并更少地应用于医学超声图像分析。

RNN也可以作为一种无监督的深度模型。在无监督的模型中，RNN通常使用以前的数据样本来预测随后的数据序列。它不需要额外的分类信息（如目标类别标签）来帮助学习，而在监督学习模型中类别序列标签是最基本的。

2.2. 无监督深度学习模型

无监督学习意味着任务特定的监督信息（如标注的类别标签）在学习过程中是不需要的。在实践应用中，利用各种无监督的深度学习模型，通过网络采样生成数据样本，如AE、RBM/DBN与广义去噪AE[64]。从这个角度来看，无监督深度模型通常作为产生式模型应用于各种任务中。下面将简要介绍医学超声图像分析中最常用的3种基本无监督深度学习模型。

2.2.1. 自编码器（AE）及其变体

简单来说，AE是一种没有涉及目标类别标签使用的非线性特征提取方法。这个方法通常用于表示学习或者在隐藏层对原始输入数据（如以输入向量的形式）进行有效的编码[9]。因此，提取的特征集中在保存与更好地表示信息上，而不是在执行特定的任务（如分类）上，尽管这两个目标并不总是互斥的。

AE通常是一个简单的网络，包含至少3层：一个输入层 x ，表示原始数据或输入特征向量（如图像中的块/像素或者语音中的频谱）；一个或多个隐藏层 h ，表示变换特征；以及一个输出层 y ，通过非线性函数来激活隐藏层从而匹配输入层 x 来完成重构：

$$h = \sigma(Wx + b) \quad (3)$$

到目前为止，已开发出许多AE的变体，如稀疏自编码器（SAE）[64]与去噪自编码器（DAE）及其堆叠版本[65]。在SAE模型中，采用正则化与稀疏化限制来增强网络训练中的求解过程，而“去噪”是防止网络学习无效解的一种解决方案。通常将AE层放置在彼此的顶部而生成这些模型的堆叠版本。

2.2.2. 限制玻尔兹曼机与深度置信网络

RBM是具有两层结构的马尔可夫随机场的一种特

殊类型[66]。就结构而言,它是一种单层无向图模型,包含一个可见层与一个隐藏层,它们之间是对称连接的,同一层中的单元之间是没有连接的。因此,RBM本质上是一种AE[67]。在实践应用中,RBM很少单独使用,一般都是一个一个堆叠起来生成更深的网络,这就成了通常的单概率模型,叫做DBM。

一个DBM包含一个可见层与若干隐藏层,最上面的两层构成一个无向二部图(如RBM),下面的层构成一个具有方向并上下连接的sigmoid置信网络。因为可以使用无标签的数据进行逐层预训练(实践中一般使用少量的标签数据),DBM具有很好的泛化能力。由于DBM使用无监督的方式进行训练,对于一个特定的任务,在实践中进行最终的微调是有必要的,通常实现监督优化的一个选项是通过在网络的最后面层增加一个线性分类器(如SVM)。对于无监督的学习模型,在最终的表示学习之后往往伴随一个微调步骤,这也是解决特定任务(如图像分类、目标检测或者组织分割)的一个最常用的实践方案。

2.3. 模型训练的困难及其应对策略

深度学习的巨大成功是源于它需要大量带有标签的训练样本才能获得优异的学习性能。然而,在当前的医学超声图像分析中,这种要求是很难满足的,这是因为专家标注是昂贵的,而且一些疾病(如病变或结节)的数据很稀有[68]。因此,在医学超声图像分析中,怎样使用有限的训练样本来训练深度模型成为了一个公开的挑战。使用有限训练样本时的最普遍的一个问题是容易产生模型过拟合。为了解决模型过拟合的问题,有两个主要的途径选择:模型优化与迁移学习。对于模型优化来说,近年来提出了很多富有成效的策略,如合理设计的初始化策略、随机梯度下降及其变体(如动量与Adagrad[69])、有效的激活函数,以及其他有力的中间正则化策略(如批量标准化),具体如下[11]:

(1) 合理设计的初始化/动量策略[70],包含合理随机初始化的利用,以及一种缓慢增加训练模型迭代过程中动量参数的特别方案。

(2) 有效的激活函数,如ReLU[71,72],一般在卷积层后面执行非线性操作。此外,Maxout也是一种激活函数[73],尤其适用于没有dropout的训练。

(3) Dropout[74]在每次训练迭代中以一定比率(如0.5)随机地使网络中的单元/神经元失活。

(4) 批量标准化[75],对训练中每个小批量数据进行标准化操作,并且在每次训练迭代中通过标准化参数反向传播梯度。

(5) 堆叠/去噪[65],主要用于AE,以便使模型变得更深,并且从损坏的输入中重建出原始“干净”的数据。

另一个关键的解决方案是迁移学习,也被广泛采用并显示了优异的性能改善能力,而且不需要大样本训练数据。这种方法避免了特定应用领域中昂贵的数据标注工作。根据Pan等[76]的研究,迁移学习可以分为三大类:归纳式迁移学习,即无论目标域与源域是否相同,目标与源任务是不同的;直推式迁移学习,即目标任务与源任务相同,而目标域与源域是不同的;以及无监督迁移学习,即归纳式迁移学习类似,只是目标任务与源任务不同,但与源任务相关。基于迁移内容,上述3类迁移学习配置所采用的方法可以分为4种情况:实例法、表示法、参数迁移法以及关系知识法。但是,本文最关心的是如何通过从其他领域(容易收集大样本训练数据的领域,如CV、语音与文本领域)迁移知识到医学超声领域来改善性能。这个过程涉及两个主要的策略:①使用预训练网络作为特征提取器(即从头开始学习特征);以及②在医学超声图像或视频序列上微调预训练的网络,这种方法广泛应用于当前的医学超声图像分析中。在一些特定的任务中,这两种策略都获得了良好的性能[77,78]。

一些其他的策略也需要注意,如数据预处理与数据增强。

2.4. 常用深度学习框架

随着相关硬件(如GPU)与软件(如开源软件库)的快速发展,深度学习技术在全世界的各种研究任务中流行起来。下面列出了5个最流行的开源深度学习框架(或软件库):

(1) Caffe[79]: <https://github.com/BVLC/caffe>;

(2) Tensorflow[80]: <https://github.com/tensorflow/tensorflow>;

(3) Theano[81]: <https://github.com/Theano/Theano>;

(4) Torch7/PyTorch[82]: <https://github.com/torch/torch7> 或 <https://github.com/pytorch/pytorch>;

(5) MXNet[83]: <https://github.com/apache/incubator-mxnet>。

大部分框架提供了多种接口,如C/C++、Matlab与Python。此外,一些软件包还提供基于这些框架写的更高级的软件库,如Keras[†]。关于这些开源框架的优缺点,请参考文献[84]。在实践应用中,研究人员可以根据实际需要以及个人喜好来选择任何框架,或者使用他们自己写的框架。

3. 深度学习在医学超声图像分析中的应用

正如前面提到的,当前深度学习技术在医学超声图像分析中的应用主要涉及三大任务:各种解剖结构(如乳腺、前列腺、肝脏、心脏与胎儿)的分类、检测与分割。本节将分别讨论每项任务在不同解剖结构中的应用。此外,在临床实践中,3D超声在改善超声成像诊断方面提供了一个有前景的方向,将作为一个独立的部分进行详细的讨论。

3.1. 分类

图像分类是诊断放射学中的一项基本认知任务,它是通过识别某些解剖或病理特征来实现的,这些特征能够区分某个解剖结构或组织与其他的解剖结构或组织。尽管当前计算机远远没有达到复现医学图像解释所需要的整个推理链,感兴趣目标(如肿瘤/病变、结节、胎儿)的自动分类是计算机辅助诊断系统中的一个研究热门。传统的机器学习方法通常利用从超声图像中提取的各种手工特征,并与多路线性分类器(如SVM)相结合来实现特定的分类任务。然而,这些方法容易受到图像失真的影响,例如由于内部或外部环境,或者成像过程中环境的影响而导致的形变。由于直接从原始数据(或者图像)中学习中层或高层抽象特征,因此深度神经网络(DNN)有一些明显的优势。此外,DN能够直接为每张图像输出一个个体预测标签,从而实现感兴趣目标的分类。对于不同的解剖应用领域,存在一些独特的挑战,下面将分别讨论。

3.1.1. 肿瘤或病变

根据疾病控制与预防中心的最新统计数据[‡],乳腺癌已成为世界各地女性中最常见的癌症和第二大癌症死亡原因。尽管乳房X线照相术仍然是临床上扫描或诊断的主要成像模式,但是超声也成为了乳腺癌诊断的重要

筛查工具。尤其,基于超声的计算机辅助诊断(CADx)系统在肿瘤疾病分类中的应用给放射医生或诊断学家提供了有效的决策支持和第二个工具选择。在传统的CADx系统中,特征提取是包括特征选择与分类在内的随后步骤的基础,它们集成在一起能够实现肿瘤或肿块病变的最终分类。乳腺或肿块病变CADx系统中的传统机器学习方法经常使用手工的和启发式的病灶提取的特征[85]。相反,深度学习能够自动地直接从图像中学习特征。

早在2012年,Jamieson等[86]进行了一项有关深度学习用于乳腺肿瘤或肿块病变分类任务的初步研究。如图4(a)所示,自适应去卷积网络(ADN)是一种无监督和产生式的分层深度模型,用于从诊断乳腺肿瘤或肿块病变超声图像中学习图像特征以及生成特征图谱,并进行了包括建立图像描述符和空间金字塔匹配(SPM)算法在内的后处理步骤。因为模型是以无监督的方式训练的,所以学习到的高层次特征(如SPM核输出)作为输入来训练监督的分类器(如线性SVM),从而实现恶性与良性乳腺肿块之间的二值分类。实验结果表明其性能已达到传统的采用人工设计特征的CADx系统的水平。在这一成功之后,许多类似的研究将深度学习方法应用于乳腺肿瘤诊断中。Liu等[87]与Shi等[19]都在两个小的超声数据集上使用了监督深度学习算法,称为深度多项式网络(DPN),或者它的堆叠版,即堆叠深度多项式网络(S-DPN)。在预处理[即基于剪切波变换的纹理特征提取和感兴趣区域(ROI)提取]和SVM分类器(或多核学习)的帮助下,获得了92.4%的最高分类准确率,优于无监督深度学习算法(如堆叠AE和DVM)。当使用深度学习从块水平的超声图像中学习图像表示,局部块不能提供丰富的上下文信息时,这种方法是一个很好的选择方案。此外,堆叠去噪自编码器(SDAE)[88]、点阵门控玻尔兹曼机(PGBM)和限制玻尔兹曼机的组合(RBM)[89],以及GoogLeNet CNN[90]也应用于乳腺超声或剪切波弹性成像来辅助乳腺癌诊断,它们都取得比人类专家更优的性能表现。在Antropova等[91]的研究中,一种融合预训练CNN提取的中低级特征与传统CADx系统获得的手工设计特征的方法应用于3个临床成像模式数据集,并证实了显著的性能改善。

另一种常见的肿瘤是肝癌,已成为全世界第六大癌

[†] <https://github.com/fchollet/keras>.

[‡] <https://www.cdc.gov/cancer/dpcp/data/women.htm>.

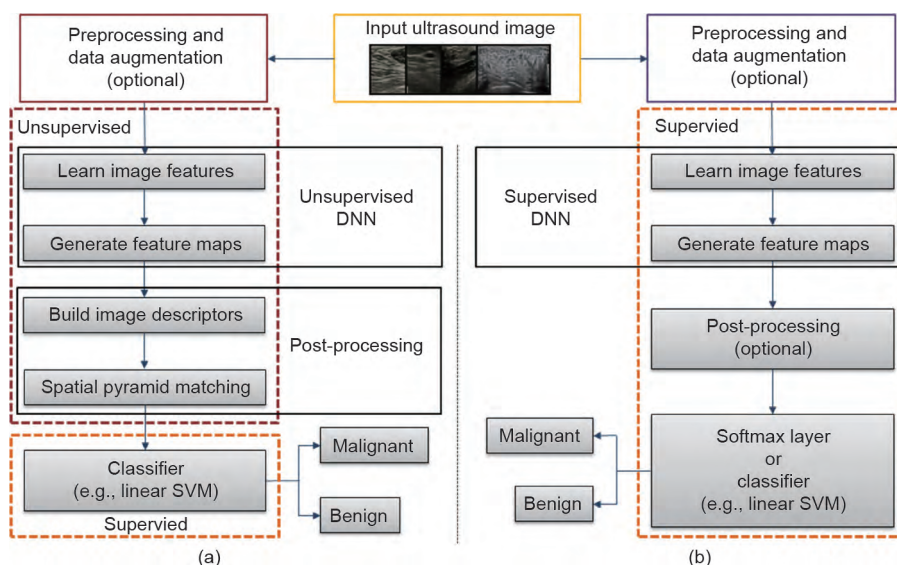


图4. 肿瘤超声图像分类流程图: (a) 无监督深度学习和 (b) 有监督深度学习。在超声图像输入到深度神经网络之前, 通常会先进行预处理和数据增强 (如提取ROI、图像裁剪等)。尽管监督深度学习中也可以使用后处理, 但是通常很少这么做, 反而会将特征图谱直接输入到softmax分类器进行分类。

症和第三大癌症死亡原因[92]。早期精准诊断对于通过提供最佳干预来提高生存率非常重要。活检仍然是当前肝癌诊断的金标准, 并且严重依赖传统CADx方法。然而, 活检是侵入性的和不舒适的, 很容易造成其他的不良影响。因此, 基于超声的诊断技术成为了肝癌检测、诊断、干预和治疗的最重要的无创方法之一。Wu等[22]将一个3层DBN应用于从对比增强超声(CEUS)视频序列中提取的时间强度曲线(TIC)中, 实现了局灶性肝脏病变的良恶性分类。他们获得了86.36%的最高准确率, 优于传统的机器学习方法[如线性判别方法(LDA)、 k 近邻(k -NN)、SVM和反向传播网络(BPN)]。为了降低使用基于TIC的特征提取方法的计算复杂度, Guo等[93]采用深度典型相关分析[DCCA, 典型相关分析(CCA)的一种变体]与多核学习分类器(MKL, 一种典型的多视角学习方法)相结合的方法来区分良恶性肝癌。实验结果表明, 充分利用这两种方法可以获得较高的分类精度(90.41%), 且有很低的计算复杂度。此外, 迁移学习策略也常应用于肝癌超声图像诊断中[58,94]。

3.1.2. 结节

甲状腺结节已成为全世界成人中最常见的结节性病变之一。当前的甲状腺结节诊断主要依靠非手术[主要是细针穿刺(FNA)活检]和手术(即切除活检)方法。然而, 这两种方法对于大规模筛查来说是很耗时耗力的, 而且可能使患者焦虑以及增加成本。随着超声技术的快速发展, 超声由于其实时性与无创特性而成为了甲

状腺结节诊断和预后的一种可选择工具。为了降低对操作者依赖和提高诊断性能, 基于超声的CADx系统易被开发出来用于检测与分类甲状腺结节。Ma等[95]在一个融合的框架里集成了两个预训练的CNN用于甲状腺结节诊断: 一个是更适合学习低级特征的浅层网络, 另一个是擅于学习高级抽象特征的深层网络。更具体地, 两个CNN在一个大的甲状腺结节超声图像数据集上分别进行训练, 然后将得到的两个特征图谱进行融合, 并输入到softmax层来诊断甲状腺结节。集成CNN学习到的高级特征与传统手工设计的低层特征也是一个可选择方案, 这已经在Liu等[96,97]的研究中得到了证实。为了克服集成特征向量中的冗余与不相关问题, 并避免过拟合, 有必要挑选特征子集。结果表明, 该方法比仅使用传统特征的方法提高了14%的精度。此外, 对于特定的任务来说, 有效的预处理和数据增强策略被证实可以提高诊断性能[48]。

3.1.3. 胎儿与新生儿

在产前超声诊断中, 胎儿生物学测量是一项必不可少的检查, 包括腹围(AC)的估计等。然而, 由于对比度低且不均匀, 以及不规则的形状, 对AC的精确测量比其他参数更加困难。在临床检查与诊断中, 不正确的胎儿AC测量可能导致胎儿体重估计不准确, 进一步增加误诊的风险[98]。因此, 胎儿超声图像的质量控制非常重要。最近, Wu等[99]提出了一个胎儿超声图像质量评价方案, 主要包含两步: ①一个CNN用于定位

ROI; ②基于ROI, 另一个CNN用于分类胎儿腹部标准面。为了改善性能, Wu等采用了一些诸如局部相位分析与图像裁剪的数据增强策略。类似地, Jang等[100]利用一个特别设计的CNN结构从超声图像中将图像块分类为关键解剖结构, 然后基于可接受的胎儿腹部平面(即标准平面), 再通过基于霍夫变换的椭圆形检测方法对胎儿AC进行估计。Gao等[101]通过胎儿解剖结构的多标签分类问题探索了从大规模自然图像到小规模超声图像学习特征的迁移性。结果表明, 迁移的CNN优于那些直接从小规模超声数据学习特征的网络(91.5% vs. 87.5%)。

胎儿心脏的定位与心脏平面的分类对于先天性心脏病的识别具有重要意义。由于胎儿心脏很小, 在临床实践中这些任务是很有挑战性的。为了解决这些问题, Sundaresan等[102]提出了将其作为语义分割问题的解决方案。更具体地, 使用一个全卷积神经网络(FCN)在超声图像帧中分割出胎儿心脏平面, 一步实现心脏的检测与心脏平面的分类。一些后处理步骤也被用于解决预测图像可能包括不同非背景区域多标签的问题。此外, Perrin等[103]在来自5个不同儿科人群的超声心动图像数据集上, 直接训练CNN来识别先天性心脏病。在一个特定的胎儿标准面识别任务中, 带有全局平均池化(GAP)策略的深层CNN, 在有限的训练数据上获得了显著的性能改善[104,105]。

3.2. 检测

在超声图像分析中, 超声图像或视频的感兴趣目标(如肿瘤、病变与结节)检测是极其重要的。尤其, 肿瘤或病变的检测可以为物体分割和良恶性肿瘤的分类提供有力的支持。解剖对象(如胎儿标准面、器官、组织或者特征点)定位也被视为分割任务或基于图像的干预与治疗的临床诊断流程的前提条件。

3.2.1. 肿瘤或病变

肿瘤/病变的检测或定位在治疗计划与干预的临床工作流程中是至关重要的, 也是最耗时耗力的任务之一。不同解剖结构的病变检测是有一些明显差异的。这项任务通常是在全图像空间上定位与识别小的病变。最近, Azizi等[20,106,107]将使用DBN从时间增强超声图像中提取的高级抽象特征与从数字病理图像中提取的组织结构相结合, 成功实现了前列腺癌的检测与分级。为了做一个全面的比较, Yap等[108]对比了3种不同的深

度学习方法: 基于块的LeNet、U-net以及迁移预训练的FCN-AlexNet, 在两个不同超声系统获得两个超声图像数据集上进行了乳腺癌检测。在两个乳腺超声图像数据集上的实验表明, 这些深度学习方法都获得了性能提高, 但是没有哪个深度学习模型在真阳性率(TPF)、每张图像的假阳性(FP)与 F 测量等指标上都获得最好的性能。类似地, Cao等[109]综合比较了4种最优的基于CNN的目标检测深度模型: Fast R-CNN [110]、Faster R-CNN [111]、You Only Look Once (YOLO) [112]与单点多框检测器(SSD) [113]用于乳腺癌检测, 结果表明SSD在精度与召回率方面都取得了最佳性能。

3.2.2. 胎儿

作为常规产科检查, 胎儿超声筛查在确认胎儿存活率、准确确定胎龄与发现影响产前护理的畸形方面发挥着至关重要的作用。在胎儿超声诊断的工作流程中, 标准面的获取是前提步骤, 对随后的生物学测量和诊断至关重要[114]。除了使用传统的机器学习方法检测胎儿超声标准面[115,116], 现在使用深度学习方法检测胎儿标准面的应用也越来越多。Baumgartner等[117,118]和Chen等[78,119]分别通过迁移的深度模型, 实现了2D超声图像中13个胎儿标准面(如肾脏、大脑、腹部、脊柱、股骨与心脏切面)和胎儿腹部(或者面部与心脏四腔心切面)标准面的检测。为了利用时空信息, 一个基于RNN的深度模型也被迁移到超声视频中多胎儿标准面(如腹部、面部水平与心脏四腔心平面)的自动化检测[60]。此外, Chen等[120]提出了基于卷积与RNN的混合通用框架, 用于超声视频中不同标准面的检测。

3.2.3. 心脏

准确识别超声心动图中心脏运动周期相位[舒张末期(ED)和收缩末期(ES)]是估计一些心脏参数(如中风量、射血分数与最大舒张容积)的必要前提条件。Dezaki等[121]提出了一种能自动识别心脏运动周期相位的深度残差循环神经网络(RRN)。RRN由残差神经网络(ResNet)、两个LSTM单元块与一个全连接层构成, 因此结合了残差神经网络和循环神经网络的优点, 前者能够处理当网络变深而出现的梯度消失或爆炸问题, 而后者能够建模序列图像帧之间的时间关系。类似地, Sofka等[122]提出了一种全卷积回归网络, 用于心脏胸骨旁长轴切面中测量点的检测, 该网络包含一个用于回归测量点位置的FCN和优化估计点位置的LSTM单元。

注意, 强化学习也与深度学习结合用于心脏超声图像的解剖关键点检测[123]。

3.3. 分割

解剖结构与病变的分割是心脏或脑分析中与体积和形状有关的临床参数的量化分析的前提条件, 在病变(如乳腺、前列腺、甲状腺结节与肺结节)的检测与分类, 以及生成ROI以便在CADx中进行后续分析方面也起着至关重要的作用。由于超声图像中目标与背景之间的对比度较低, 对大部分解剖结构, 尤其病变(结节)超声图像进行准确的分割, 仍然是一个很有挑战性的任务。此外, 众所周知, 手动分割方法是非常耗时耗力的, 且存在很大的个体差异。因此, 很有必要开发更先进的自动分割方法来解决这些问题。使用深度学习进行解剖结构分割的一些结果如图5所示[21,38,44,46,50,57,124–126]。

3.3.1. 非刚性器官

超声心动图因其成本较低, 可用性与可移植性强, 而成为可视化与诊断心脏左心室(LV)最常用的成像模态之一。为了诊断心脏病, 心脏病学家必须对心脏进行定量功能分析, 这通常需要对收缩末期与舒张末期的LV进行准确的分割。显然, LV的手动分割非常繁琐、耗时且具有主观性, 而自动LV分割系统可能有潜力解决这些问题。然而, 由于显著的外形与形状差异、低信噪比、阴影以及边缘不完整, 全自动分割LV仍然是一项具有挑战性的任务。为了解决这些问题, 各种传统机器学习方法, 如主动轮廓[127]和可变形模板[128], 已被广泛用于LV的自动化分割, 这通常需要使用与LV形状和外形有关的先验知识。最近, 基于深度学习的方法也被频繁采用。Carneiro等[129–134]利用能够从原始超声图像中学习高级特征的DNN来自动地分割LV。为了提高性能, 还采用了一些其他的策略, 如高效搜索方法、

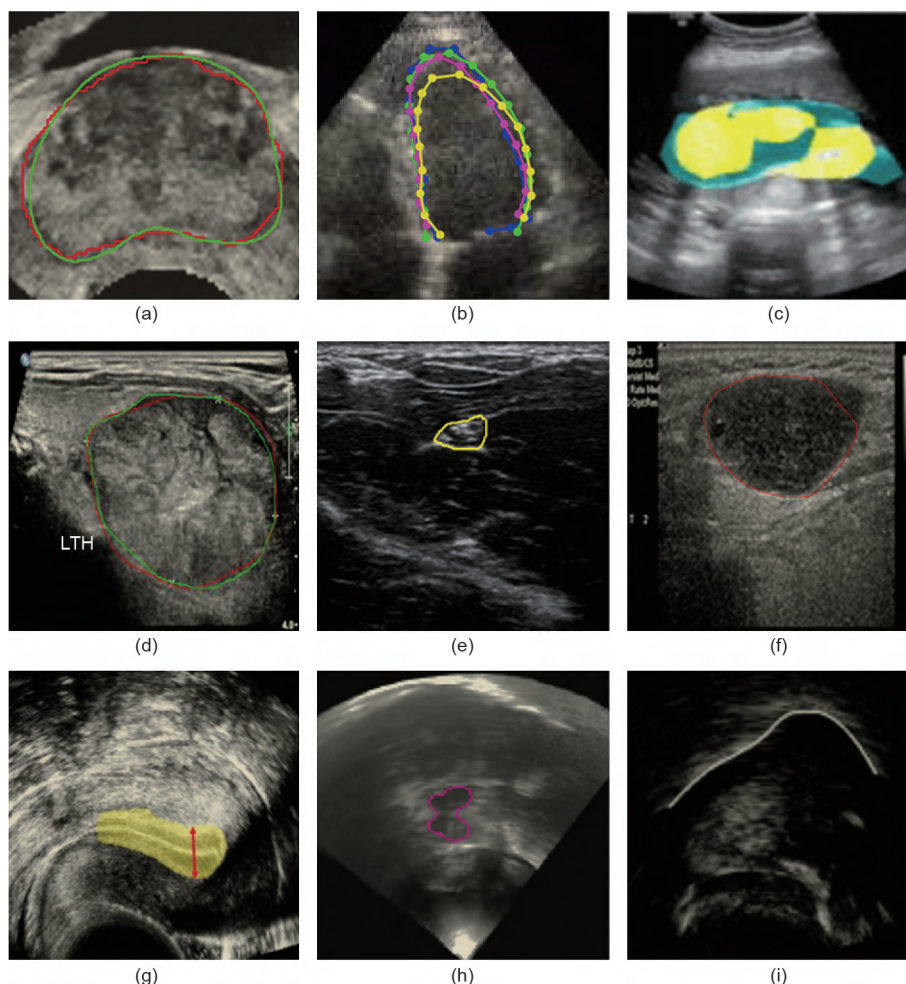


图5. 使用深度学习进行常见解剖结构分割的分割结果示例。(a) 前列腺[21]; (b) 左心室[124]; (c) 羊水与胎儿身体[50]; (d) 甲状腺结节[125]; (e) 中神经结构[44]; (f) 淋巴结[38]; (g) 子宫内层[126]; (h) 中脑[57]; (i) 舌轮廓[46]。所有这些结果都表明了与人类放射科医生相当的分割性能。不同颜色的线或点表示相应的分割轮廓或区域。

粒子滤波器、在线协同训练方法以及多重动态模型。

典型的非刚性分割方法通常把分割问题分为两步：①刚性检测与②非刚性分割或轮廓描绘。第一步非常重要，因为它可以减少搜索时间以及降低训练复杂度。为了降低刚性检测中的训练与推理复杂度，同时保持较高的分割精度，Nascimento和Carneiro[124,135]利用稀疏流形学习方法结合DBN来进行非刚性物体的分割。实验结果表明，在刚性检测中使用稀疏流形学习与DBN的结合取得了与当前最优结果一样精确的性能，但是具有更低的训练与搜索复杂度。不像典型的非刚性分割方法，Nascimento和Carneiro[136]通过显式轮廓的稀疏低维流形映射直接进行非刚性分割，但是泛化能力有限。尽管大部分研究表明，使用深度学习能够取得比传统机器学习方法更好的性能，但是最近的一项研究[137]表明，在2D超声心动图像的LV分割中，手工制作特征优于CNN，且训练中的计算成本更低。一个合理的解释是，用于手工设计特征的监督下降（SDM）[138]回归方法在迭代优化估计的LV轮廓方面更加灵活。

与成人LV分割相比，胎儿LV分割更具有挑战性，这是因为胎儿超声心动图像序列存在不均匀性、伪影、对比度差和被试间差异太大等特性；此外，因为胎儿在子宫内随机运动，LV与左心房（LA）通常是连接在一起的。为了解决这些问题，Yu等[139]提出了一种基于多尺度信息与微调的动态CNN方法，用于胎儿LV分割。该动态CNN对每一个超声心动图像序列的第一帧和余下帧分别进行深微调和浅微调，以便适应每个胎儿。此外，还采用了匹配的方法来分离LV与LA之间的连接区域。实验表明，与固定的CNN相比，动态CNN在平均Dice系数上取得了从88.35%到94.5%的显著性能提高。

3.3.2. 刚性器官

在医学超声图像中，许多解剖结构/对象（如前列腺、乳腺、肾脏、胎儿等）的边界不完整是一个普遍问题，这对这些结构的自动分割提出了很大的挑战。目前，有两种主要的方法来解决这个问题：①自下而上法，在监督下将每个像素分为前景（目标）或者背景；②自上而下法，利用先验形状信息来引导分割。通过端到端的、全监督学习的方式对图像中的每一个像素进行分类，许多研究实现了对不同解剖结构（如胎儿身体与羊水[50]、淋巴结[38]以及骨骼[140]）进行像素级分割的任务。对于特定的任务，这些研究中提出的深度学习方法在性能与速度方面都优于最先进的方法。

自下而上法的一个显著优势是，它能为图像中每个像素提供预测；但是，由于缺少先验形状信息，它可能无法处理边界信息丢失的问题。相比之下，自上而下法可以通过对形状建模为分割任务提供强有力的形状引导，尽管合适的形状建模通常是困难的。为了同时实现关键点描述符学习与形状推理，Yang等[21]将边界完整性表述为一个序列问题，即动态建模形状。为了同时利用自下而上法与自上而下法，Ravishankar等[39]利用先前从形状正则网络中学习到形状，来优化FCN分割网络获得的预测分割结果。在一个肾脏超声数据上的实验结果表明，先验形状信息的利用可以使肾脏分割的性能提高大约5%。此外，Wu等[141]将FCN网络核植入到自动上下文模型[142]以便利用局部上下文信息，从而解决了严重边界不完整的问题，且显著提高了分割精度。Anas等[143]在基于残差网络（ResNet）的深度框架优化中，使用指数加权图来提高局部预测。

解决分割任务的另一种方法是将分割问题表述为块级分类问题，正如文献[125]所述。该方法可以显著降低计算成本与内存需求。

3.4. 3D 超声图像分析

由于3D深度学习的应用困难，目前应用于医学超声图像分析的深度学习方法大多应用在2D图像，尽管输入可能是3D的。事实上，由于以下限制，3D深度学习仍然是一项具有挑战性的任务：①在大体积数据上训练一个深度网络，对于实际的临床应用来说，计算成本可能太高（如内存与计算要求显著增加）；②以3D图像块作为输入的深度网络需要更多的训练样本，这是由于3D网络包含的参数与2D网络相比呈指数级增加。在有限训练数据的情况下，这可能显著增加过拟合的风险[144]。相反，由于很难生成与共享病变或疾病图像，医学超声图像分析领域常常只能收集到有限的训练样本（即使在使用了数据增强后，样本数通常也只有几百或者几千份）。然而，在医学超声图像分析领域，越来越多的人试图解决这些具有挑战性的3D深度学习任务。

在常规妇科超声检查和绝经后出血的女性子宫内膜癌筛查中，通常采用厚度测量法进行子宫内膜的评估。Singhal等[126]提出了一种基于FCN的两步算法，实现子宫内膜厚度的全自动测量。首先，提出了一种混合变分曲线-传播模型，叫做深度学习蛇形（DLS）分割模型，用于从3D经阴道超声体积数据中检测和分割子宫内膜。该模型将深度学习的子宫内膜概率图集成到分割能量函

数中, 而该概率图是在矢状切面上通过基于U-net的子
宫内膜定位来预测构建的。分割后, 以分割掩膜中两个
接触面(基底层)之间的最大距离作为测量的厚度。

为了解决产科和慢性疼痛治疗中, 超声引导的硬膜
外针注射时针状目标的自动定位问题, Pesteic等[145]提
出了一种卷积网络结构以及特征增强技术。该方法包含
两步: ①利用局部有向Hadamard (LDH) 特征与前向
反馈神经网络从3D超声体积数据中分类平面; ②在识
别的目标平面中, 通过CNN分类图像中的每个像素而实
现目标定位。

Nie等[146]提出了一种从复杂3D超声数据中自动检
测中矢状面的方法。为了避免不必要的大规模搜索以及
相应的巨大计算负载, 他们巧妙地将矢状面检测问题转
化为对称面与轴搜索问题。更具体地说, 该方法包含3
步: ①根据文献[147], 建立一个DBN, 从3D超声数据
的中间切面中检测完全包含胎儿头部的图像块; ②采用
增强的圆检测方法, 确定胎儿头部在图像块中的位置与
大小; ③最后, 通过一个模型以及前两步确定的胎儿头
部位置与大小等先验知识, 确定矢状面。

应该指出的是, 这3个方法实际上都是基于2D深度
学习的逐片处理(slice-by-slice)方法, 尽管它们都可
以用于3D超声体积数据。这样做的优点是高速、低内
存消耗, 并且能够直接或通过迁移学习利用预训练的网
络。然而, 缺点是无法利用图像平面正交方向的结构上
下文信息。为了克服这些缺点, Milletari等[57]提出了一
种称为霍夫-CNN (Hough-CNN) 的逐块多图谱方法,
用于多个深部脑区域的检测与分割。该方法采用了类似
于早期研究中提出的霍夫投票策略, 区别在于结构特定
的特征是通过CNN而不是SAE获得的。为了充分利用
3D超声体积数据中的上下文信息, Pourtaherian等[148]
直接训练了一个3D CNN来检测3D超声体积数据中的针
状体素, 每个体素是通过从以它为中心的3个正交面中
局部提取的原始数据来分类的。为了解决数据集中高度
不平衡的问题, 采用了一种新的更新策略被采用, 即在
训练阶段对非针状体素进行重采样, 从而提高检测性能
以及鲁棒性。

广泛应用于2D超声图像的典型非刚性物体分割方
案也适用于3D超声体积的分割。Ghesu等[52]采用这种
典型的非刚性分割方法, 实现了3D超声体积主动脉瓣
的分割, 该方法包含两步: 刚性目标定位与非刚性目
标边界估计。为了解决3D目标检测的问题, 采用了边
缘空间深度学习方法 (MSDL), 该方法结合了边缘空

间学习 (MSL) [149]和深度学习。基于检测到的目标,
对非刚性形状做一个初始估计, 然后采用基于稀疏自
适应DNN的主动形状模型来引导形状变形。在一个大
型3D经食管超声心动图像数据集上的实验结果证实了,
MSDL在主动脉瓣的3D检测与分割任务中的效率与鲁棒
性, 它的性能比当前最先进的方法提高了42.5%。仅使
用中央处理单元 (CPU), 主动脉瓣可以在不到1 s的时
间内成功分割, 其准确度高于原来的MSL。

胎儿结构的分割比一般的解剖结构或器官更具挑战
性。例如, 胎盘高度可变, 因为它的位置取决于子宫内
的着床部位。虽然已证实手动分割与半自动分割方法是
准确的以及可接受的, 但是它们特别耗时且依赖于操
作者的经验。为了解决这些问题, Looney等[150]采用
DeepMedic从3D超声体积中分割胎盘。训练数据集没有
使用人工标注数据, 而是使用半自动随机步行 (RW)
方法的输出作为标注结果。DeepMedic是由Kamnitsas等
[151]提出的一种双通道3D CNN结构, 最开始用于大脑
磁共振成像 (MRI) 数据的病变分割。然而, 3D超声
体积的胎盘成功分割表明, DeepMedic是一个适用于不
同模态3D医学体积数据的通用3D深度学习结构。最近,
Yang等[152]将RNN植入到定制的3D FCN中, 用于同时
分割超声体积中的多个目标, 包括胎儿、妊娠囊与胎盘。
针对普遍存在的边界不确定性问题, 采用了一种有效的
序列化策略。此外, 他们还提出了一种分层深监督机制,
以促进循环神经网络中的信息流动, 进一步提高分割性
能。类似地, Schmidt-Richberg等[153]结合了FCN与可
变形形状模型, 用于3D胎儿腹部超声体积的分割。

4. 挑战与应用前景

从上面的例子可以看出, 深度学习已经应用到医学
超声图像分析中的各种应用领域。然而, 尽管深度学习
方法不断更新医学超声图像分析中不同应用方面的最新
性能结果, 但仍有改进的空间。本节将总结深度学习应
用于医学超声图像分析时普遍遇到的挑战, 并讨论它的
未来发展前景。

显然, 深度学习能取得的主要性能改善, 在很大
程度上依赖于大样本训练数据集。然而, 与其他领域的
大规模与公共可用的数据集 (如ImageNet数据集超过
 1×10^6 标注的多类别自然图像[6]) 相比, 当前医学超
声领域内公开可用的数据集仍然很有限。有限的数据集
已成为深度学习方法在医学超声图像分析中进一步

应用的瓶颈。

为解决小样本数据集的问题，目前研究人员最常用的方法之一是进行跨数据集（模态内或模态间）学习，即迁移学习。正如前文所述，迁移学习的使用主要分为两类：直接利用预训练网络作为特征提取器以及固定网络中的部分权重进行微调[77]。根据目标域与源域是否相同，迁移学习可以分为两类：跨模态与跨领域迁移学习。跨领域迁移学习是医学超声图像分析中各任务最常用的方法。在任何情况下，当前的模型预训练总是在大样本数据集上进行。这样做可以确保出色的性能，但这在医学成像领域绝对不是最佳选择。当使用小样本训练数据集时，特定领域深度模型从头开始训练（如果能够恰当选择模型大小）能够比其他领域大样本数据集（如自然图像）预训练网络后的迁移学习取得更好的性能[154]。这一现象的根本原因可能是，用于医学图像分析中特定任务的原始输入图像像素到特征向量的映射，在预训练情况下变得更加复杂，因而需要大样本训练数据集才能有好的泛化能力。相反，特别设计的小网络可能更适合医学成像中普遍存在的小规模训练数据集[155]。因此，开发医学成像领域专用的深度学习模型，不仅能够以较低计算复杂度来提高特定任务的性能，还可以促进医学成像领域中CADx的技术发展。

此外，在自然图像上训练的模型可能不是医学图像的最佳模型，后者通常是单通道、低对比度以及具有丰富的纹理特征。在医学成像，尤其乳腺成像中，诊断工作流程中经常使用多种成像模态，如MRI、X射线与超声。超声或者乳房X线照相术通常是一线筛查工具，因此更容易收集到大量的训练数据集。然而，通常用于筛查高风险人群的乳腺MRI是一种更加昂贵和耗时的方法，要收集足够的训练数据集以及进行标注将更加困难。在这种情况下，跨模态迁移学习可能是一个可取的选择。有实验表明，在缺乏足够训练数据集的情况下，对于特定的任务，跨模态迁移学习要优于跨领域迁移学习[156]。因为考虑到很少能够从单个站点（即机构或医院）收集到大样本数据集，而通常收集自多个不同的站点（或机器），因此可以尝试进行相同成像模态的跨站点（或跨机器）迁移学习。

最后，还必须解决当前迁移学习算法的其他问题，包括如何避免负性迁移，如何处理源域与目标域或者源任务与目标任务之间的特征空间不一致问题，以及如何提高不同任务的泛化能力。迁移学习的目的是利用从源任务学习到的知识来提高目标任务的学习性能。

然而，不恰当的迁移学习有时反而会降低性能，即成为负迁移[157]。

忽略不同方法间的固有差异，对于特定的目标任务，任何迁移学习的有效性主要取决于两个方面：源任务及其与目标的关系。理想情况下，迁移方法将在足够相关的任务间产生正迁移，而避免负迁移，尽管这些任务可能不是完全匹配的。然而，这些目标在实践中很难同时实现。为了避免负迁移，可以采用以下策略：①识别与拒绝有害源任务知识；②从一组候选源任务中选择最好的源任务（如果可能的话）；③建模多个候选源任务的相似性。此外，当源任务和目标任务表示不一致时，映射可以用于任务表示之间的迁移。

值得再次强调的是，3D超声是医学成像领域中一种重要的成像模态，3D超声图像分析在基于超声的临床应用中显示出巨大的潜力，尽管一些问题亟待解决。可以预见的是，更多新颖的3D深度学习算法将被开发，用于医学超声图像分析中的各种任务，并在未来实现更大的性能改进。然而，如果没有其他领域，尤其CV的支持，当前医学超声图像分析领域的3D深度学习算法开发将比较困难。

致谢

本工作由国家自然科学基金（61571304、81571758和61701312）、国家重点研发计划（2016YFC0104703）、广东省医学科研基金（B2018031）和深圳市孔雀计划（KQTD2016053112051497）资助。

Compliance with ethics guidelines

Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2018.11.020>.

References

- [1] Reddy UM, Filly RA, Copel JA. Prenatal imaging: ultrasonography and

- magnetic resonance imaging. *Obstet Gynecol* 2008;112(1):145–57.
- [2] Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging* 2006;25(8):987–1010.
 - [3] Salomon LJ, Winer N, Bernard JP, Ville Y. A score-based method for quality control of fetal images at routine second-trimester ultrasound examination. *Prenat Diagn* 2008;28(9):822–7.
 - [4] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6):84–90.
 - [5] Wang G. A perspective on deep imaging. *IEEE Access* 2016;4:8914–24.
 - [6] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;115(3):211–52.
 - [7] Salakhutdinov R. Learning deep generative models. *Annu Rev Stat Appl* 2015;2(1):361–85.
 - [8] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
 - [9] Deng L, Yu D. Deep learning: methods and applications. *Found Trends Signal Process* 2014;7(3–4):197–387.
 - [10] Deng L, Li J, Huang JT, Yao K, Yu D, Seide F, et al. Recent advances in deep learning for speech research at Microsoft. In: *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2013 May 26–31; Vancouver, BC, Canada. New York: IEEE; 2013. p. 8604–8.
 - [11] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19(1):221–48.
 - [12] Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 2016;35(5):1153–9.
 - [13] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
 - [14] Suzuki K. Overview of deep learning in medical imaging. *Radiological Phys Technol* 2017;10(3):257–73.
 - [15] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access* 2018;6:9375–89.
 - [16] Anas EMA, Seitel A, Rasoulia A, John PS, Pichora D, Darras K, et al. Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. *Int J CARS* 2015;10(6):959–69.
 - [17] Hiramatsu Y, Muramatsu C, Kobayashi H, Hara T, Fujita H. Automated detection of masses on whole breast volume ultrasound scanner: false positive reduction using deep convolutional neural network. In: *Proceedings of the SPIE Medical Imaging*; 2017 Feb 11–16; Orlando, FL, USA. Bellingham: SPIE; 2017.
 - [18] Bian C, Lee R, Chou Y, Cheng J. Boundary regularized convolutional neural network for layer parsing of breast anatomy in automated whole breast ultrasound. In: *Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, editors. Medical image computing and computer-assisted intervention—MICCAI 2017*. Berlin: Springer; 2017. p. 259–66.
 - [19] Shi J, Zhou S, Liu X, Zhang Q, Lu M, Wang T. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. *Neurocomputing* 2016;194:87–94.
 - [20] Azizi S, Imani F, Zhuang B, Tahmasebi A, Kwak JT, Xu S, et al. Ultrasoundbased detection of prostate cancer using automatic feature selection with deep belief networks. In: *Navab N, Hornegger J, Wells W, Frangi A, editors. Medical image computing and computer-assisted intervention—MICCAI 2015*. Berlin: Springer; 2015. p. 70–7.
 - [21] Yang X, Yu L, Wu L, Wang Y, Ni D, Qin J, et al. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*; 2017 Feb 4–9; San Francisco, CA, USA: AAAI Press; 2017. p. 1633–9.
 - [22] Wu K, Chen X, Ding M. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik* 2014;125(15):4057–63.
 - [23] Ghesu FC, Georgescu B, Zheng Y, Hornegger J, Comaniciu D. Marginal space deep learning: efficient architecture for detection in volumetric image data. In: *Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical image computing and computer-assisted intervention*. Berlin: Springer; 2015. p. 710–8.
 - [24] Pereira F, Bueno A, Rodriguez A, Perrin D, Marx G, Cardinale M, et al. Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms. *J Med Imaging* 2017;4(1):014502.
 - [25] Sombune P, Phienphanich P, Phuechpanpaisal S, Muengtawepong S, Ruamthanthong A, Tantibundhit C. Automated embolic signal detection using deep convolutional neural network. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway: IEEE; 2017. p. 3365–8.
 - [26] Milletari F, Ahmadi SA, Kroll C, Hennemersperger C, Tombari C, Shah A, et al. Robust segmentation of various anatomies in 3D ultrasound using hough forests and learned data representations. In: *Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical image computing and computer-assisted intervention*. Berlin: Springer; 2015. p. 111–8.
 - [27] Lekadir K, Galimzianova A, Betriu A, Del Mar Vila M, Igual L, Rubin DL, et al. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound. *IEEE J Biomed Health Inform* 2017;21(1):48–55.
 - [28] Shin J, Tajbakhsh N, Hurst RT, Kendall CB, Liang J. Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 2526–35.
 - [29] Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Med Phys* 2017;44(5):1678–91.
 - [30] Smistad E, Løvtakken L. Vessel detection in ultrasound images using deep convolutional neural networks. In: *Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep learning and data labeling for medical applications*. Berlin: Springer; 2016. p. 30–8.
 - [31] Su S, Gao Z, Zhang H, Lin Q, Hao WK, Li S. Detection of lumen and mediaadventitia borders in IVUS images using sparse auto-encoder neural network. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21, Melbourne, Australia. Piscataway: IEEE; 2017. p. 1120–4.
 - [32] Yaqub M, Kelly B, Papageorgiou AT, Noble JA. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. *Ultrasound Med Biol* 2017;43(12):2925–33.
 - [33] Huang W, Bridge CP, Noble JA, Zisserman A. Temporal HeartNet: towards human-level automatic analysis of fetal cardiac screening video. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Berlin: Springer; 2017. p. 341–9.
 - [34] Gao Y, Noble JA. Detection and characterization of the fetal heartbeat in freehand ultrasound sweeps with weakly-supervised two-streams convolutional networks. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Berlin: Springer; 2017. p. 305–13.
 - [35] Qi H, Collins S, Noble A. Weakly supervised learning of placental ultrasound images with residual networks. In: *Hernández MV, González-Castro V, editors. Medical image understanding and analysis*. Berlin: Springer; 2017. p. 98–108.
 - [36] Chen H, Zheng Y, Park JH, Heng PA, Zhou K. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In: *Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2016 Oct 17–21; Athens, Greece. Berlin: Springer; 2016. p. 487–95.
 - [37] Ravishankar H, Prabhu SM, Vaidya V, Singhal N. Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning. In: *Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging*; 2016 Jun 13–16; Prague, Czech Republic. Piscataway: IEEE; 2016. p. 779–82.
 - [38] Zhang Y, Ying MTC, Yang L, Ahuja AT, Chen DZ, et al. Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: *Proceedings of 2016 IEEE International Conference on Bioinformatics and Biomedicine*; 2016 Dec 15–18; Shenzhen, China. Piscataway: IEEE; 2016. p. 443–8.
 - [39] Ravishankar H, Venkataramani R, Thiruvankadam S, Sudhakar P, Vaidya V. Learning and incorporating shape models for semantic segmentation. In: *Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention*; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 203–11.
 - [40] Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R. SLIDE: automatic spine level identification system using a deep convolutional neural network. *Int J CARS* 2017;12(7):1189–98.
 - [41] Golan D, Donner Y, Mansi C, Jaremko J, Ramachandran M. Fully automating Graf's method for DDH diagnosis using deep convolutional neural networks. In: *Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep learning and data labeling for medical applications*. Proceedings of International Workshops on DLMI and LABELS; 2016 Oct 21; Athens, Greece. Berlin: Springer; 2016. p. 130–41.
 - [42] Hareendranathan AR, Zonoobi D, Mabee M, Cobzas D, Punithakumar K, Noga ML, et al. Toward automatic diagnosis of hip dysplasia from 2D ultrasound. In: *Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging*; 2017 Apr 18–21; Melbourne, Australia. 2017. p. 982–5.
 - [43] Burlina P, Billings S, Joshi N, Albayda J. Automated diagnosis of myositis from muscle ultrasound: exploring the use of machine learning and deep learning methods. *PLoS ONE* 2017;12(8):e0184059.
 - [44] Hafiane A, Vieyres P, Delbos A. Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images. 2017. arXiv:1706.05870.
 - [45] Fasel I, Berry J. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In: *Proceedings of 2010 20th International Conference on Pattern Recognition*; 2010 Aug 23–26; Istanbul, Turkey. p. 1493–6.
 - [46] Jaumard-Hakoun A, Xu K, Roussel-Ragot P, Dreyfus G, Denby B. Tongue contour extraction from ultrasound images based on deep neural network. 2016. arXiv:1605.05912.
 - [47] Xu K, Roussel P, Csapó TG, Denby B. Convolutional neural network-based automatic classification of midsagittal tongue gestural targets using B-mode ultrasound images. *J Acoust Soc Am* 2017;141(6):EL531–7.
 - [48] Chi J, Wallia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging* 2017;30(4):477–86.
 - [49] Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J Digit Imaging* 2017;30

- (2):234–43.
- [50] Li Y, Xu R, Ohya J, Iwata H. Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers. In: Proceedings of 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2017 Jul 11–15; Seogwipo, Korea. Piscataway: IEEE; 2017. p. 1485–8.
 - [51] Fenster A, Downey DB, Cardinal HN. Three-dimensional ultrasound imaging. *Phys Med Biol* 2001;46(5):R67–99.
 - [52] Ghesu FC, Krubasik E, Georgescu B, Singh V, Zheng Y, Horneegger J, et al. Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE Trans Med Imaging* 2016;35(5):1217–28.
 - [53] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 2017;30(4):449–59.
 - [54] Xing F, Xie Y, Su H, Liu F, Yang L. Deep learning in microscopy image analysis: a survey. *IEEE Trans Neural Networks Learn Syst* 2017;29(10):4550–68.
 - [55] Xian M, Zhang Y, Cheng HD, Xu F, Zhang B, Ding J. Automatic breast ultrasound image segmentation: a survey. *Pattern Recognit* 2018;79:340–55.
 - [56] He K, Sun J. Convolutional neural networks at constrained time cost. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition; 2015 Oct 7–12; Boston, MA, USA. Piscataway: IEEE; 2015. p. 5353–60.
 - [57] Milletari F, Ahmadi SA, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput Vis Image Underst* 2017;164:92–102.
 - [58] Liu X, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors* 2017;17(1):149.
 - [59] Canziani A, Paszke A, Culurciello E. An analysis of deep neural network models for practical applications. 2016. arXiv:1605.07678.
 - [60] Chen H, Dou Q, Ni D, Cheng J, Qin J, Li S, et al. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: Proceedings of 2015 IEEE Medical Image Computing and Computer-Assisted Intervention; 2015 Oct 5–9; Munich, Germany. Berlin: Springer; 2015. p. 507–14.
 - [61] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 1994;5(2):157–66.
 - [62] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
 - [63] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014. arXiv:1406.1078.
 - [64] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013;35(8):1798–828.
 - [65] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 2010;11(12):3371–408.
 - [66] Hinton GE. A practical guide to training restricted boltzmann machines. In: Montavon G, Orr GB, Müller KR, editors. *Neural networks: tricks of the trade*. Berlin: Springer; 2012. p. 599–619.
 - [67] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313(5786):504–7.
 - [68] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 2016;35(5):1299–312.
 - [69] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 2011;12(7):2121–59.
 - [70] Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on International Conference on Machine Learning; 2013 Jun 16–21; Atlanta, GA, USA. JMLR; 2013. p. 1139–47.
 - [71] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning; 2010 Jun 21–24; Haifa, Israel. Piscataway: Omnipress; 2010. p. 807–14.
 - [72] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics; 2011 Apr 11–13; Ft. Lauderdale, FL, USA. JMLR; 2011. p. 315–23.
 - [73] Goodfellow IJ, Warde-Farley D, Mirza M, Courville A, Bengio Y. Maxout networks. In: Proceedings of the 30th International Conference on Machine Learning; 2013 Jun 16–21; Atlanta, GA, USA. JMLR; 2013. p. 1319–27.
 - [74] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. 2012. arXiv: 1207.0580.
 - [75] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning; 2015 Jul 6–11; Lille, France. JMLR; 2015. p. 448–56.
 - [76] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345–59.
 - [77] Azizi S, Mousavi P, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection. *Int J CARS* 2017;12(7):1111–21.
 - [78] Chen H, Ni D, Qin J, Li S, Yang X, Wang T, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Health Inform* 2015;19(5):1627–36.
 - [79] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia; 2014 Nov 3–7; New York, NY, USA. New York: ACM; 2014. p. 675–8.
 - [80] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. 2016. arXiv: 1603.04467.
 - [81] Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow L, Bergeron A, et al. Theano: new features and speed improvements. 2012. arXiv: 1211.5590.
 - [82] Collobert R, Kavukcuoglu K, Farabet C. Torch7: a matlab-like environment for machine learning. In: Proceedings of the NIPS 2011 Workshop; 2011 Dec 12–17; Granada, Spain, 2011.
 - [83] Chen T, Li M, Li Y, Lin M, Wang N, Wang M. MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. 2015. arXiv: 1512.01274.
 - [84] Bahrampour S, Ramakrishnan N, Schott L, Shah M. Comparative study of deep learning software frameworks. 2015. arXiv: 1511.06435.
 - [85] Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med Phys* 2008;35(12):5799–820.
 - [86] Jamieson A, Drukker K, Giger M. Breast image feature learning with adaptive deconvolutional networks. In: Proceedings of the SPIE Medical Imaging; 2012 Feb 4–9; San Diego, CA, USA. Bellingham: SPIE; 2012.
 - [87] Liu X, Shi J, Zhang Q. Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset. In: Proceedings of the 6th International Workshop on Machine Learning in Medical Imaging; 2015 Oct 5; Munich, Germany. Berlin: Springer; 2015. p. 313–20.
 - [88] Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep* 2016;6(1):24454.
 - [89] Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* 2016;72:150–7.
 - [90] Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol* 2017;62(19):7714–28.
 - [91] Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys* 2017;44(10):5162–71.
 - [92] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127(12):2893–917.
 - [93] Guo L, Wang D, Xu H, Qian Y, Wang C, Zheng X, et al. CEUS-based classification of liver tumors with deep canonical correlation analysis and multi-kernel learning. In: Proceedings of 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2017 Jul 11–15; Seogwipo, Korea. Piscataway: IEEE; 2017. p. 1748–51.
 - [94] Meng D, Zhang L, Cao G, Cao W, Zhang G, Hu B. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access* 2017;5:5804–10.
 - [95] Ma J, Wu F, Zhu J, Xu D, Kong D. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* 2017;73:221–30.
 - [96] Liu T, Xie S, Yu J, Niu L, Sun WD. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. In: Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing; 2017 Jun 19; New Orleans, LA, USA. Piscataway: IEEE; 2017. p. 919–23.
 - [97] Liu T, Xie S, Zhang Y, Yu J, Niu L, Sun W. Feature selection and thyroid nodule classification using transfer learning. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 1096–9.
 - [98] Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 2002;19(2):190–6.
 - [99] Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. *IEEE Trans Cybern* 2017;47(5):1336–49.
 - [100] Jang J, Kwon JY, Kim B, Lee SM, Park Y, Seo JK. CNN-based estimation of abdominal circumference from ultrasound images. 2017. arXiv: 1702.02741.
 - [101] Gao Y, Maraci MA, Noble JA. Describing ultrasound video content using deep convolutional neural networks. In: Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging. Piscataway: IEEE; 2016. p. 787–90.
 - [102] Sundaresan V, Bridge CP, Ioannou C, Noble A. Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 671–4.
 - [103] Perrin DP, Bueno A, Rodriguez A, Marx GR, Del Nido PJ. Application of convolutional artificial neural networks to echocardiograms for differentiating congenital heart diseases in a pediatric population. In: Proceedings of the SPIE Medical imaging 2017: Computer-aided Diagnosis; 2017 Mar 3; Orlando, FL, USA. Bellingham: SPIE; 2012.
 - [104] Yu Z, Ni D, Chen S, Li S, Wang T, Lei B. Fetal facial standard plane recognition

- via very deep convolutional networks. In: Proceedings of 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2016 Aug 16–20; Orlando, FL, USA. Piscataway: IEEE; 2016. p. 627–30.
- [105] Yu Z, Tan EL, Ni D, Qin J, Chen S, Li S, et al. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. *IEEE J Biomed Health Inform* 2018;22(3):874–85.
- [106] Azizi S, Imani F, Ghavidel S, Tahmasebi A, Kwak JT, Xu S, et al. Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. *Int J CARS* 2016;11(6):947–56.
- [107] Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. *Int J CARS* 2017;12(8):1293–305.
- [108] Yap MH, Pons G, Martí J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* 2018;22(4):1218–26.
- [109] Cao Z, Duan L, Yang G, Yue T, Chen Q, Fu C, et al. Breast tumor detection in ultrasound images using deep learning. In: Wu G, Munsell B, Zhan Y, Bai W, Sanroma G, Coupé P, editors. Patch-based techniques in medical imaging. Berlin: Springer; 2017. p. 121–8.
- [110] Girshick R. Fast R-CNN. In: Proceedings of 2015 IEEE International Conference on Computer Vision; 2015 Dec 7–13; Santiago, Chile. Piscataway: IEEE; 2015. p. 1440–8.
- [111] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39(6):1137–49.
- [112] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, realtime object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 779–88.
- [113] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, et al. SSD: single shot multibox detector. In: Proceedings of the European Conference on Computer Vision; 2016 Oct 11–14; Amsterdam, The Netherlands. Berlin: Springer; 2016. p. 21–37.
- [114] D. Ni, X. Yang, X. Chen, et al. Standard Plane Localization in Ultrasound by Radial Component Model and Selective Search. *Ultrasound in Medicine and Biology*, 2014, 40(11): 2728–2742.
- [115] Yang X, Ni D, Qin J, Li S, Wang T, Chen S, et al. Standard plane localization in ultrasound by radial component. In: Proceedings of 2014 IEEE 11th International Symposium on Biomedical Imaging; 2014 Apr 29–May 2; Beijing, China. Piscataway: IEEE; 2014. p. 1180–3.
- [116] Ni D, Li T, Yang X, Qin J, Li S, Chin C, et al. Selective search and sequential detection for standard plane localization in ultrasound. In: Yoshida H, Warfield S, Vannier MW, editors. Abdominal imaging, computation and clinical applications. Berlin: Springer; 2013. p. 203–11.
- [117] Baumgartner CF, Kamnitsas K, Matthew J, Smith S, Kainz B, Rueckert D, et al. Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17–21; Athens, Greece. Berlin: Spring; 2016. p. 203–11.
- [118] Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans Med Imaging* 2017;36(11):2204–15.
- [119] Chen H, Ni D, Yang X, Li S, Heng PA. Fetal abdominal standard plane localization through representation learning with knowledge transfer. In: Wu G, Zhang D, Zhou L, editors. Machine learning in medical imaging. Berlin: Springer; 2014. p. 125–32.
- [120] Chen H, Wu L, Dou Q, Qin J, Li S, Cheng JZ, et al. Ultrasound standard plane detection using a composite neural network framework. *IEEE Trans Cybern* 2017;47(6):1576–86.
- [121] Dezaki FT, Dhungel N, Abdi A, Luong C, Tsang T, Jue J, et al. Deep residual recurrent neural networks for characterization of cardiac cycle phase from echocardiograms. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, editors. Deep learning in medical image analysis and multimodal learning for clinical decision support. Berlin: Springer; 2017. p. 100–8.
- [122] Sofka M, Milletari F, Jia J, Rothberg A. Fully convolutional regression network for accurate detection of measurement points. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, editors. Deep learning in medical image analysis and multimodal learning for clinical decision support. Berlin: Springer; 2017. p. 258–66.
- [123] Ghesu FC, Georgescu B, Mansi T, Neumann D, Hornegger J, Comaniciu D. An artificial agent for anatomical landmark detection in medical images. In: Proceedings of 2016 IEEE Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17–21; Athens, Greece. Berlin: Springer; 2016. p. 229–37.
- [124] Nascimento JC, Carneiro G. Multi-atlas segmentation using manifold learning with deep belief networks. In: Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging; 2016 Apr 13–16; Prague, Czech Republic. Piscataway: IEEE; 2016. p. 867–71.
- [125] Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int J CARS* 2017;12(11):1895–910.
- [126] Singhal N, Mukherjee S, Perrey C. Automated assessment of endometrium from transvaginal ultrasound using Deep Learned Snake. In: Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 83–6.
- [127] Bernard O, Touil B, Gelas A, Prost R, Friboulet D. A RBF-Based multiphase level set method for segmentation in echocardiography using the statistics of the radiofrequency signal. In: Proceedings of 2007 IEEE International Conference on Image Processing; 2007 Oct 16–19; San Antonio, TX, USA. Piscataway: IEEE; 2007.
- [128] Jacob G, Noble JA, Behrenbruch C, Kelion AD, Banning AP. A shape-spacebased approach to tracking myocardial borders and quantifying regional leftventricular function applied in echocardiography. *IEEE Trans Med Imaging* 2002;21(3):226–38.
- [129] Carneiro G, Nascimento J, Freitas A. Robust left ventricle segmentation from ultrasound data using deep neural networks and efficient search methods. In: Proceedings of 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2010 Apr 14–17; Rotterdam, The Netherlands. Piscataway: IEEE; 2010. p. 1085–8.
- [130] Carneiro G, Nascimento JC. Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures. In: Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010 Jun 13–18. San Francisco, CA, USA. Piscataway: IEEE; 2010. p. 2815–22.
- [131] Carneiro G, Nascimento JC. Incremental on-line semi-supervised learning for segmenting the left ventricle of the heart from ultrasound data. In: Proceedings of 2011 International Conference on Computer Vision; 2011 Nov 6–13; Barcelona, Spain. Piscataway: IEEE; 2011. p. 1700–7.
- [132] Carneiro G, Nascimento JC. The use of on-line co-training to reduce the training set size in pattern recognition methods: application to left ventricle segmentation in ultrasound. In: Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16–21; Providence, RI, USA. Piscataway: IEEE; 2012. p. 948–55.
- [133] Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Trans Image Process* 2012; 21(3):968–82.
- [134] Carneiro G, Nascimento JC. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Trans Pattern Anal Mach Intell* 2013;35(11):2592–607.
- [135] Nascimento JC, Carneiro G. Deep learning on sparse manifolds for faster object segmentation. *IEEE Trans Image Process* 2017;26(10):4978–90.
- [136] Nascimento JC, Carneiro G. Non-rigid segmentation using sparse low dimensional manifolds and deep belief networks. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. Piscataway: IEEE; 2010. p. 288–95.
- [137] Raynaud C, Langet H, Amzulescu MS, Saloux E, Bertrand H, Allain P, et al. Handcrafted features vs. ConvNets in 2D echocardiographic images. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 1116–9.
- [138] Xiong X, Torre FDL. Global supervised descent method. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition; 2015 Jun 7–12; Boston, MA, USA. Piscataway: IEEE; 2015. p. 2664–73.
- [139] Yu L, Guo Y, Wang Y, Yu J, Chen P. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. *IEEE Trans Biomed Eng* 2017;64(8):1886–95.
- [140] Baka N, Leenstra S, van Walsum T. Ultrasound aided vertebral level localization for lumbar surgery. *IEEE Trans Med Imaging* 2017;36(10):2138–47.
- [141] Wu L, Xin Y, Li S, Wang T, Heng P, Ni D. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 663–6.
- [142] Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2010;32(10):1744–57.
- [143] Anas EMA, Nouranian S, Mahdavi SS, Spadinger I, Morris WJ, Salcudean SE, et al. Clinical target-volume delineation in prostate brachytherapy using residual neural networks. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 365–73.
- [144] Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D. 3D deep learning for efficient and robust landmark detection in volumetric data. In: Proceedings of 2015 IEEE Medical Image Computing and Computer-Assisted Intervention; 2015 Oct 5–9; Munich, Germany. Piscataway: IEEE; 2015. p. 565–72.
- [145] Pesteie M, Lessoway V, Abolmaesumi P, Rohling RN. Automatic localization of the needle target for ultrasound-guided epidural injections. *IEEE Trans Med Imaging* 2018;37(1):81–92.
- [146] Nie S, Yu J, Chen P, Wang Y, Zhang JQ. Automatic detection of standard sagittal plane in the first trimester of pregnancy using 3-D ultrasound data. *Ultrasound Med Biol* 2017;43(1):286–300.
- [147] Nie S, Yu J, Chen P, Zhang J, Wang Y. A novel method with a deep network and directional edges for automatic detection of a fetal head. In: Proceedings

- of 2015 the 23rd European Signal Processing Conference; 2015 Aug 31–Sep 4; Nice, France. Piscataway: IEEE; 2015. p. 654–8.
- [148] Pourtaherian A, Zanjani FG, Zinger S, Mihajlovic N, Ng G, Korsten H, et al. Improving needle detection in 3D ultrasound using orthogonal-plane convolutional networks. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 610–8.
- [149] Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans Med Imaging* 2008;27(11):1668–81.
- [150] Looney P, Stevenson GN, Nicolaides KH, Plasencia W, Molloy M, Natsis S, et al. Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 279–82.
- [151] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61–78.
- [152] Yang X, Yu L, Li S, Wang X, Wang N, Qin J, et al. Towards automatic semantic segmentation in volumetric ultrasound. In: Proceedings of 2017 IEEE Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11–13; Quebec City, Canada. Piscataway: IEEE; 2017. p. 711–9.
- [153] Schmidt-Richberg A, Brosch T, Schadeewaldt N, Klinder T, Caballaro A, Salim I, et al. Abdomen segmentation in 3D fetal ultrasound using CNN-powered deformable models. In: Proceedings of the 4th International Workshop on Fetal and Infant Image Analysis; 2017 Sep 14; Quebec City, Canada. Piscataway: IEEE; 2017. p. 52–61.
- [154] Amit G, Ben-Ari R, Hadad O, Monovich E, Granot N, Hashoul S. Classification of breast MRI lesions using small-size training sets: comparison of deep learning approaches. In: Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis; 2017 Mar 3; Orlando, Florida. Bellingham: SPIE; 2017.
- [155] Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35(5):1285–98.
- [156] Hadad O, Bakalo R, Ben-Ari R, Hashoul S, Amit G. Classification of breast lesions using cross-modal deep learning. In: Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Apr 18–21; Melbourne, Australia. Piscataway: IEEE; 2017. p. 109–12.
- [157] Lisa T, Jude S. Transfer learning. In: Olivas ES, Guerrero JDM, Martinez-Sober M, Magdalena-Benedito JR, López AJ, editors. Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. Hershey: IGI Global; 2010. p. 242–64.