



Research  
AI for Precision Medicine—Review

## Deep Learning in Medical Ultrasound Analysis: A Review

Shengfeng Liu<sup>a</sup>, Yi Wang<sup>a</sup>, Xin Yang<sup>b</sup>, Baiying Lei<sup>a</sup>, Li Liu<sup>a</sup>, Shawn Xiang Li<sup>a</sup>, Dong Ni<sup>a,\*</sup>, Tianfu Wang<sup>a,\*</sup>



<sup>a</sup> 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

<sup>b</sup> 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](mailto:nidong@szu.edu.cn) (D. Ni), [tfwang@szu.edu.cn](mailto:tfwang@szu.edu.cn) (T. Wang).

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 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].

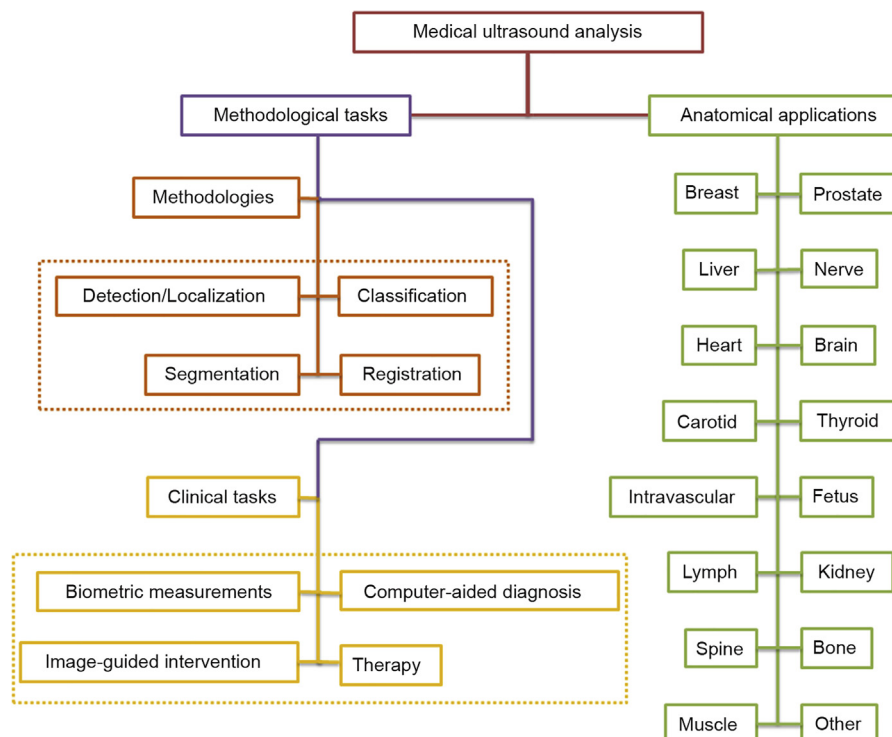


Fig. 1. Illustration of medical US analysis.

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.

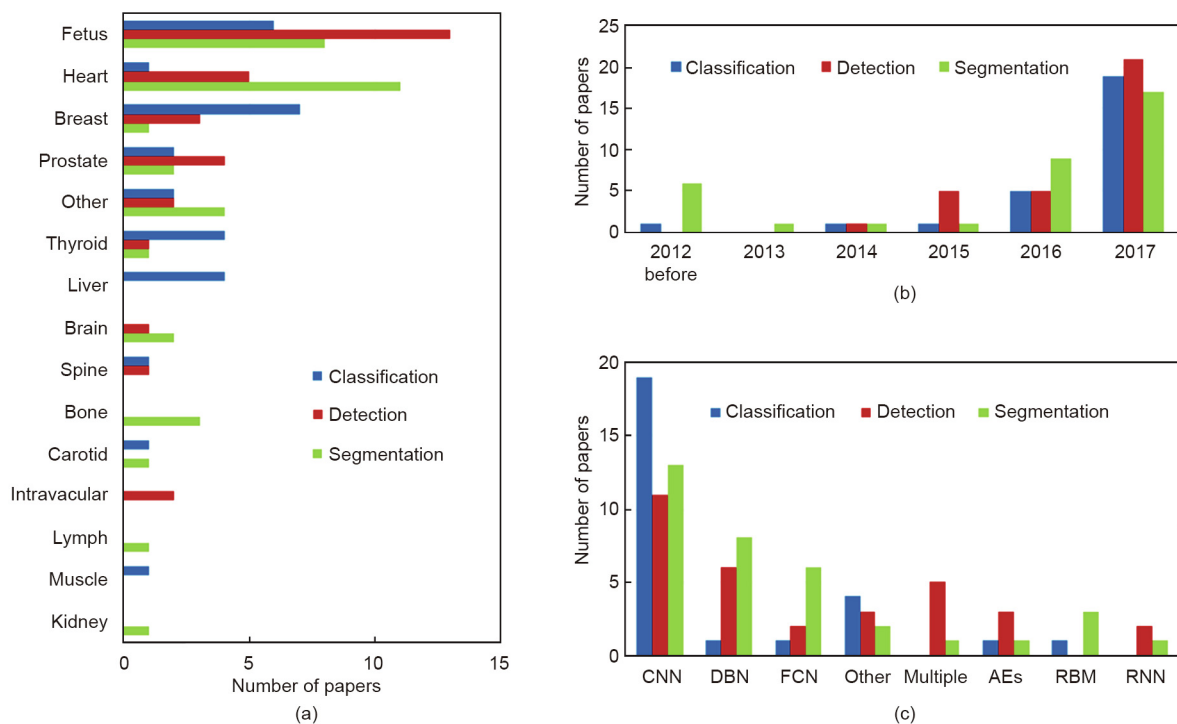
### 2.1. Supervised deep models

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.

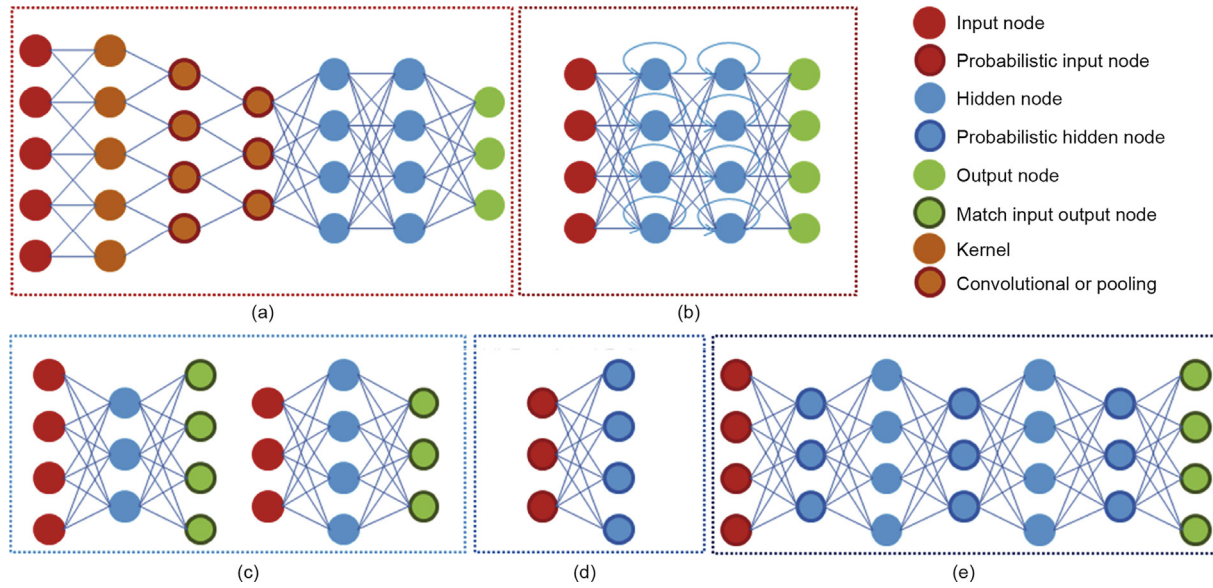
#### 2.1.1. Convolutional neural networks

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/SAE (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  $\sigma$  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<sup>†</sup>. 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.

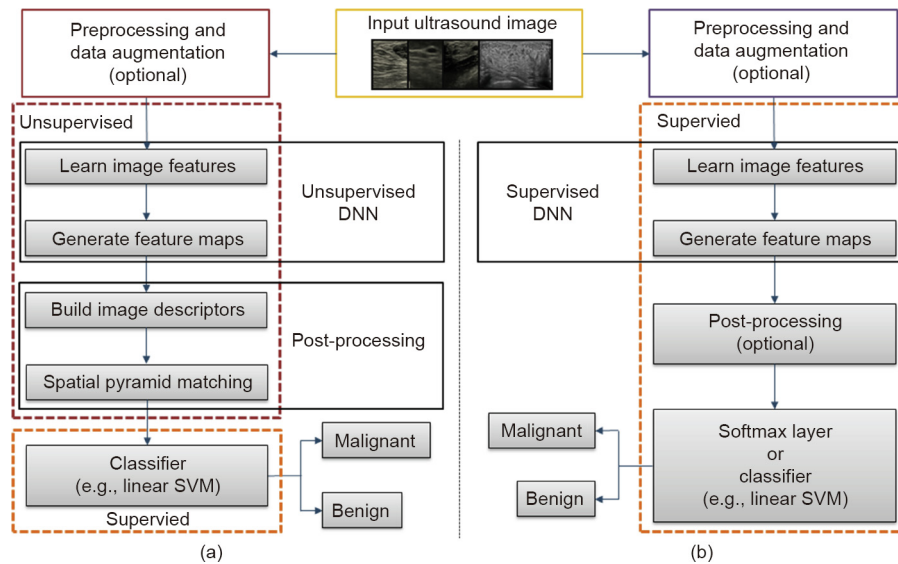
##### 3.1.1. Tumors or lesions

According to the latest statistics from the Centers for Disease Control and Prevention<sup>‡</sup>, 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.

<sup>†</sup> <https://github.com/fchollet/keras>.

<sup>‡</sup> <https://www.cdc.gov/cancer/dcp/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].

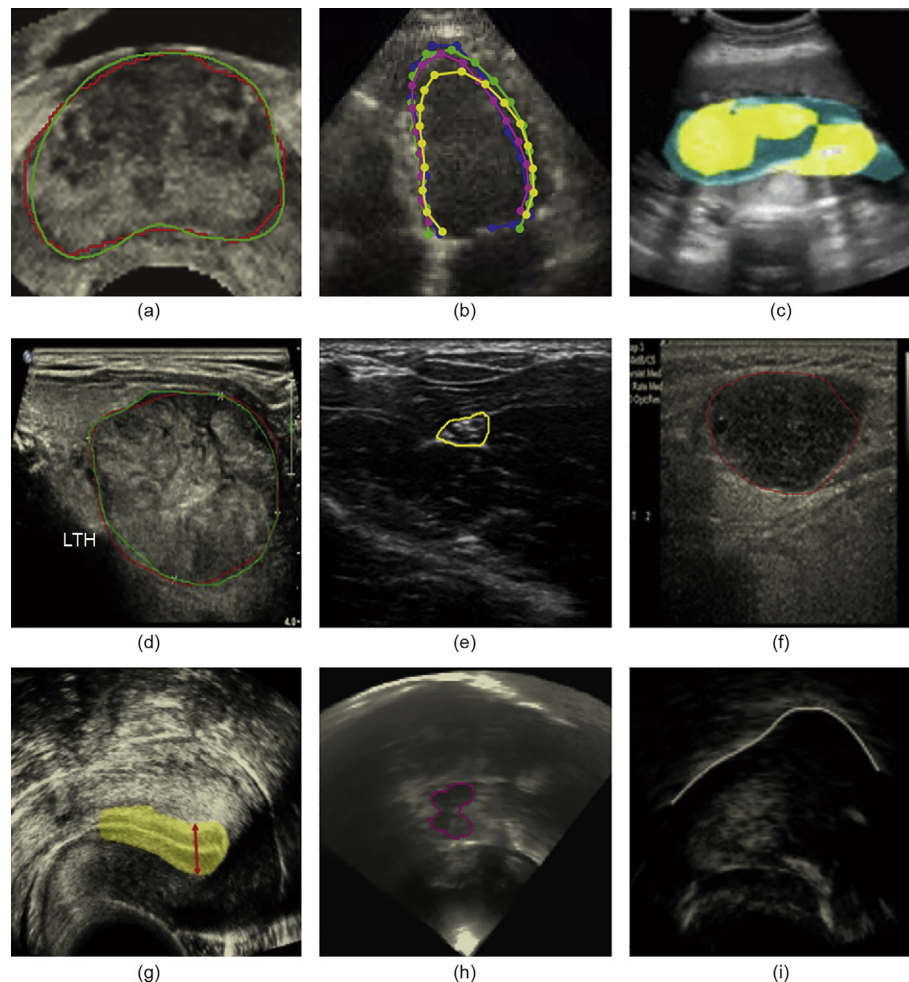
### 3.3. Segmentation

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, 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. 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, 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, Løvestakken 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] 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 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 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 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 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, 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, 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, Angelov 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  $\text{Na}^+/\text{Ca}^{2+}$  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, Scheuring 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, Nicolaidis 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, 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.