FUIQA: Fetal Ultrasound Image Quality Assessment With Deep Convolutional Networks

Lingyun Wu†, Jie-Zhi Cheng†, *Member, IEEE*, Shengli Li, Baiying Lei, *Member, IEEE*, Tianfu Wang, and Dong Ni*

Abstract—The quality of ultrasound (US) images for the obstetric examination is crucial for accurate biometric measurement. However, manual quality control is a labor intensive process and often impractical in a clinical setting. To improve the efficiency of examination and alleviate the measurement error caused by improper US scanning operation and slice selection, a computerized fetal US image quality assessment (FUIQA) scheme is proposed to assist the implementation of US image quality control in the clinical obstetric examination. The proposed FUIQA is realized with two deep convolutional neural network models, which are denoted as L-CNN and C-CNN, respectively. The L-CNN aims to find the region of interest (ROI) of the fetal abdominal region in the US image. Based on the ROI found by the L-CNN, the C-CNN evaluates the image quality by assessing the goodness of depiction for the key structures of stomach bubble and umbilical vein. To further boost the performance of the L-CNN, we augment the input sources of the neural network with the local phase features along with the original US data. It will be shown that the heterogeneous input sources will help to improve the performance of the L-CNN. The performance of the proposed FUIQA is compared with the subjective image quality evaluation results from three medical doctors. With comprehensive experiments, it will be illustrated that the computerized assessment with our FUIQA scheme can be comparable to the subjective ratings from medical doctors.

Index Terms—Deep convolutional neural network (DCNN), fetal ultrasound (US), local phase, quality control.

I. INTRODUCTION

ULTRASOUND (US) imaging is a widely used obstetric examination tool, and the biometric measurements like
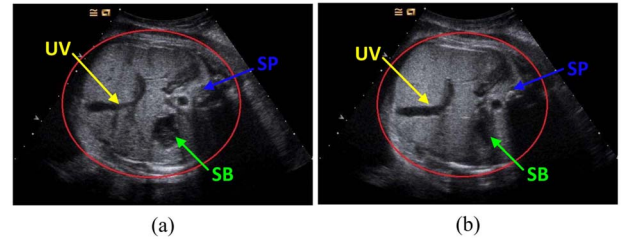
Fig. 1. Illustration of fetal abdominal US image. (a) Standard plane. (b) NG US image.

fetal weight can be further estimated for assessing fetal health. The fetal US images for measurements are commonly termed as standard planes. Fig. 1 illustrates examples of an US standard plane and a not good (NG) US image, which were acquired from the same fetus at the same day but with different sonographers. The images shown in Fig. 1 were for fetal abdominal circumference (AC) measurement. It can be found that the stomach bubble (SB) was better pictured with clear boundary in the US standard plane. The AC measurements (red circles in Fig. 1) from the standard plane and NG image were 338.72 and 326.87 mm, respectively. The difference is 11.85 mm, simply because of the NG quality.

In general, the development of fetal AC is around 10 mm per week at the first trimester. This means the AC measurement error around 5–10 mm could lead to significantly incorrect estimation of fetal weight. The incorrect weight estimation may increase the risk of misdiagnosis [1]. When AC is too small, it may refer to intrauterine growth restriction, which is one of the major causes for the 60% fetus deaths [2]. On the other hand, when AC exceeds the normal range at the first trimester, it may suggest that the baby grows too much and the chance of caesarean would increase [3]. Referring to Fig. 1, the measurement error with NG image can be up to the unacceptable range of 5–10 mm. In such case, inaccurate AC measurement may result in serious misdiagnosis. Therefore the quality control for the acquired US images is very crucial for the obstetric examination.

In a number of studies [1], [4], [5], the importance of quality control for fetal US imaging has been highlighted. Particularly in [5], a quality control guideline was developed for the subjective evaluation of the US fetal images. Although the guideline defined in [5] was shown to be effective, there may exist two potential drawbacks with the subjective assessment. First, the assessment may be highly biased with evaluator's experience

WU et al.: FUIQA WITH DEEP CONVOLUTIONAL NETWORKS

TABLE I
QUALITY ASSESSMENT CRITERIA FOR FETAL ABDOMINAL US IMAGES. AN US IMAGE WILL BE SCORED
WITH ONE POINT WHENEVER A CRITERION IS MET

Item	Criteria	Score
ROI	The area of ROI shall be larger than the half of overall fan-shaped area of US view.	1
SB	The fetal stomach bubble (SB) shall appear full and salient with clear boundary. The interior of SB is expected to have hypo-echogenicity.	1
UV	The fetal umbilical vein (UV) shall appear in curved shape without disconnection. The interior of UV may sometimes have medium echogenicity.	1

and suffers from interobserver variation. Second, since the assessment process involves repetitive and laborious manual operations, the clinical implementation of the quality control protocol will require extra workload and hence may not be practical. To address the two drawbacks, we propose an automatic quality assessment scheme to attain the goals of: 1) objective evaluation; 2) robustness to the variation of imaging conditions and image contents; and 3) more efficiency. The fetal US image quality assessment (FUIQA) scheme is abbreviated as FUIQA for convenience. We specifically apply the FUIQA scheme on the fetal abdominal images with the evaluation criteria in Table I. These criteria are defined with the reference of guideline [5] and further revised by our clinical advisor committee at the Shenzhen Maternal and Child Healthcare Hospital, Shenzhen, China. The committee is led by a senior radiologist with more than 20 years of experience in the fetal US examination. As an example, the standard plane shown in Fig. 1(a) appears to meet all criteria in Table I.

The automatic FUIQA scheme developed here can perform the complete quality assessment of fetal US images with the output of evaluation score. To our best knowledge, there is very little relevant study on this topic. Previous studies only did partial job or targeted on different directions. For examples, Rahmatullah et al. [6], [7] attempted to exploit the computerized quality assessment for the fetal US images, but only stopped at the detection of the SB and umbilical vein (UV) structures without the evaluation of goodness of depiction for these structures as defined in Table I. Meanwhile, the methods developed in [6] and [7] are semiautomatic. The work [8] proposed a method to automatically retrieve the standard planes from the fetal US videos. The method developed a radial component model to describe the relative geometric relation among the structures of fetal spine (SP), UV, and SB for the avoidance of misidentification of gall bladder (GB) as UV. To bypass the difficult problems of feature extraction, Chen et al. [9], [10] further exploited the deep learning techniques with the aid of transfer learning frameworks. Although promising performances had been reported in [9] and [10] for standard plane retrieval, the methods [9], [10] cannot be directly applied to our problem as the goal is different.

The development of the FUIQA scheme may encounter several challenges, which are summarized in Fig. 2. The first challenge is the low quality of US images [Fig. 2(a)], whereas the second one is the significant variation in size and appearance of the fetal abdominal region [Fig. 2(b)]. The third challenge is the appearance similarity of different fetal structures and high appearance variations of the same structures [see Fig. 2(c)], where the UV and GB are quite similar. The fourth challenge is the gap between the computational and semantic domains [11]–[13]. Referring to Table I, the
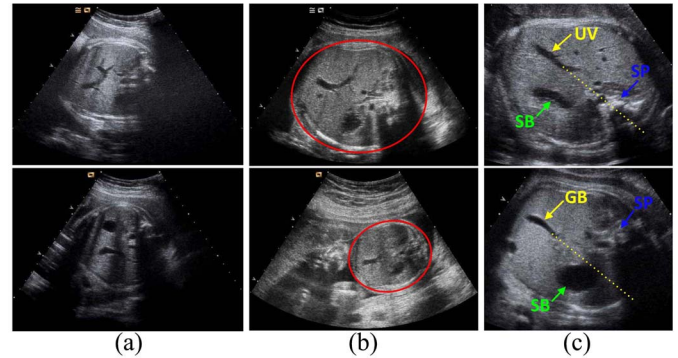

Fig. 2. Cases that illustrate different types of challenges. (a) US images with serious noise and shadowing effect. (b) Cases of fetal abdominal regions (marked with red circles) with significant variations in size and appearance. (c) US images with similar but different fetal structures in relatively similar positions. In (c), the differentiation of these two cases can be based on the geometric relations between the structures by yellow and blue arrows. To identify UV structure, the global direction of UV shall direct to the SP's center whereas the direction of GB does not.

descriptions of the SB and UV in the second and third criteria are put in a literal way. In particular, the literal descriptions like "full and salient," "curve-shaped," "with clear boundary," etc., are hard to be quantified and can be subjective.

The quantification of each criterion in Table I requires the elaboration of finding useful image features. Since useful image features for each criterion are unknown, we here employ the deep learning technique for its advantage of automatic feature discovery [14], [15]. Deep learning is an emerging technique that can successfully address many medical image analysis problems like differentiation of pulmonary nodules and breast lesions [16], [17], abnormality detections [18]–[20], prediction and identification of brain diseases [21], structure segmentation [22]–[25], standard plane retrieval [9], [10], biometric measurement [26], etc. The proposed FUIQA scheme is based on the deep convolutional neural network (DCNN) for automatically realizing the criteria in Table I. A point can be scored whenever an US image satisfies a criterion in Table I. A fetal abdominal US image will be regarded as a standard plane and scored in 3 points if it meets all criteria in Table I. To implement the first criterion, a DCNN model, called L-CNN, is constructed to identify the region of interest (ROI) of the fetal abdominal region in the US image. Based on the L-CNN, another DCNN, denoted as C-CNN, is further built to jointly evaluate the SB and UV structures on the identified ROI. The C-CNN aims to implement the criteria 2 and 3 with its capability of automatic modeling of the relation among the structures of SB, UV, and SP, as well as the shape/appearance variation with respect to the SB and UV structures. The L-CNN and C-CNN models are cast as classification schemes. The L-CNN aims to identify good ROIs from the whole US

images, whereas the C-CNN model attempts to categorize the ROIs into 4 classes of all possible cases of the goodness of depiction for SB and UV structures.

To better adapt the DCNN architecture into the context of FUIQA, two implementation issues are elaborated. The two issues are whether to: 1) incorporate the data augmentation procedure [27] and 2) import the redundant channel inputs with local phase features [28], [29] in the DCNN. Since the DCNN architecture is commonly equipped with three input channels, most DCNN working for medical image analysis simply replicate the image data for all channels [9], [10]. In this paper, we further investigate the usage of local phase symmetric and asymmetric features as the alternative input sources. The local phase features were shown to be helpful in several US image analysis applications [7], [30]–[32]. The contributions of this paper can be threefold. First, a new computerized FUIQA is proposed to assist the quality control of the US image acquisition. The new FUIQA scheme may potentially alleviate the workload of tedious image review process and render the quality control more feasible in clinical practice. To our best knowledge, this is the first comprehensive FUIQA work. The second contribution is the comprehensive study on the two implementation issues of DCNN for FUIQA. Particularly, we conducted a rigorous survey on the usage of local phase features for the L-CNN and C-CNN architectures. To our best knowledge, there is little work that explores the usage of heterogeneous input sources of DCNN. Third, the proposed FUIQA scheme can be easily generalized to other types of fetal US views for the depiction of fetal face, fetal four cardiac chambers, etc. The quality assessment for the other view types also includes the steps of ROI localization and goodness of depiction for the relevant key structures. Provided with the annotations of other views, the FUIQA can be easily applied.

II. METHOD

The flowchart of our FUIQA scheme is illustrated in Fig. 3. With the heterogeneous input sources of original US data, symmetric and asymmetric phase features, the L-CNN is able to localize the fetal abdominal ROI. The C-CNN then further analyzes the identified ROI with the classification scheme. The C-CNN rates the input US image in Fig. 3 with the score of 2, since the UV in the image is not well depicted. In this section, we will briefly introduce the DCNN and then elaborate the details of the L-CNN training, the ROI localization, the C-CNN training, and the fetal abdominal ROI classification. Afterward, the two implementation issues will be discussed.

A. Deep Convolutional Neural Network

The DCNN model is a popular deep learning technique and has been successfully applied to address many medical image analysis problems [9], [10], [18]–[20]. It is a biologically inspired neural network model, which is able to exploit the spatial patterns from the training image data in a layer-by-layer fashion with the convolutional kernels. Provided with sufficiently large training data and proper parameter setting,

the DCNN can address many complicated pattern recognition problems. A typical DCNN model can be constituted of pairs of convolutional (C) and max-pooling (M) layers and commonly ended with fully connected layers on the top of the network architecture. Several feature maps can be obtained by convolving the learnable linear filters at convolutional layer with the previous layer (input image for the first convolutional layer). Specifically, given that the h_j^l is the jth feature map in the lth layer and h_n^{l-1} $(n = 1, \ldots, N)$ is the nth feature map in the $(l-1)$th layer, the feature map h_j^l can be computed by

$$h_j^l = \sigma \left(\sum_{n=1}^{N} \left(W_{jn}^l * h_n^{l-1} \right) + b^l \right) \qquad (1)$$

where W_{jn}^l is the convolutional kernel (linear filter) connected to the nth feature map of the previous layer and b^l is the bias term for the lth layer. $\sigma(\cdot)$ is the nonlinear neural activation function of the rectified linear unit (ReLu) and can be defined as $\sigma(x) = \max(x, 0)$. The ReLu was shown to be capable of creating sparse representation with hard linearity and may help to attain better classification performance than the conventional sigmoid function does [33].

The M layers perform suppression of nonmaximal values to offload the computation for the latter layers and shall be robust to translation effect. The fully connected layer, where neurons are fully connected, is the conventional neural network model of multilayer perceptron. The fully connected layer aims to conduct high level reasoning and nonlinear feature combinations. For the classification purpose, a DCNN model can be ended with soft-max functions at the final output layer. In the training process, the internal parameters can be adjusted with the minimization of the loss function of all training samples x_i $(i = 1, \ldots, N)$. The loss function can be defined as

$$\ell(x_i, y_i; \theta) = -\sum_{i}^{N} \sum_{k=1}^{K} \left(1\{y_i = k\} \log \left[p(c_k = 1|x_i, \theta) \right] \right) \qquad (2)$$

$$p(c_k = 1|x_i, \theta) = \frac{e^{o_k}}{\sum_{j=1}^{K} e^{o_j}}, \quad k = 1, \ldots, K \qquad (3)$$

where $p(c_k = 1|x_i, \theta)$ is the probability of classifying the training sample x_i as the kth class of the overall K classes. θ represents the set of internal parameters to be estimated, whereas y_i denotes the class label of the training sample x_i. $1\{\cdot\}$ is the indicator function and $1\{y_i = k\}$ suggests y_i belongs to the kth class with the value 1; otherwise, it will be 0. o_k can be taken as the high level output features for the input of the soft-max layer. The minimization of the cost function $\ell(x_i, y_i; \theta)$ can be achieved by the technique of stochastic gradient descending with the fashion of minibatch training.

B. ROI Localization With the L-CNN

The L-CNN aims to identify the ROI that encloses the fetal abdominal region from the US image. It is formulated as a two-class differentiation scheme. For achieving better architecture initialization, the knowledge encoded in the AlexNet [27] is transferred to the L-CNN to reuse the low-level cues from
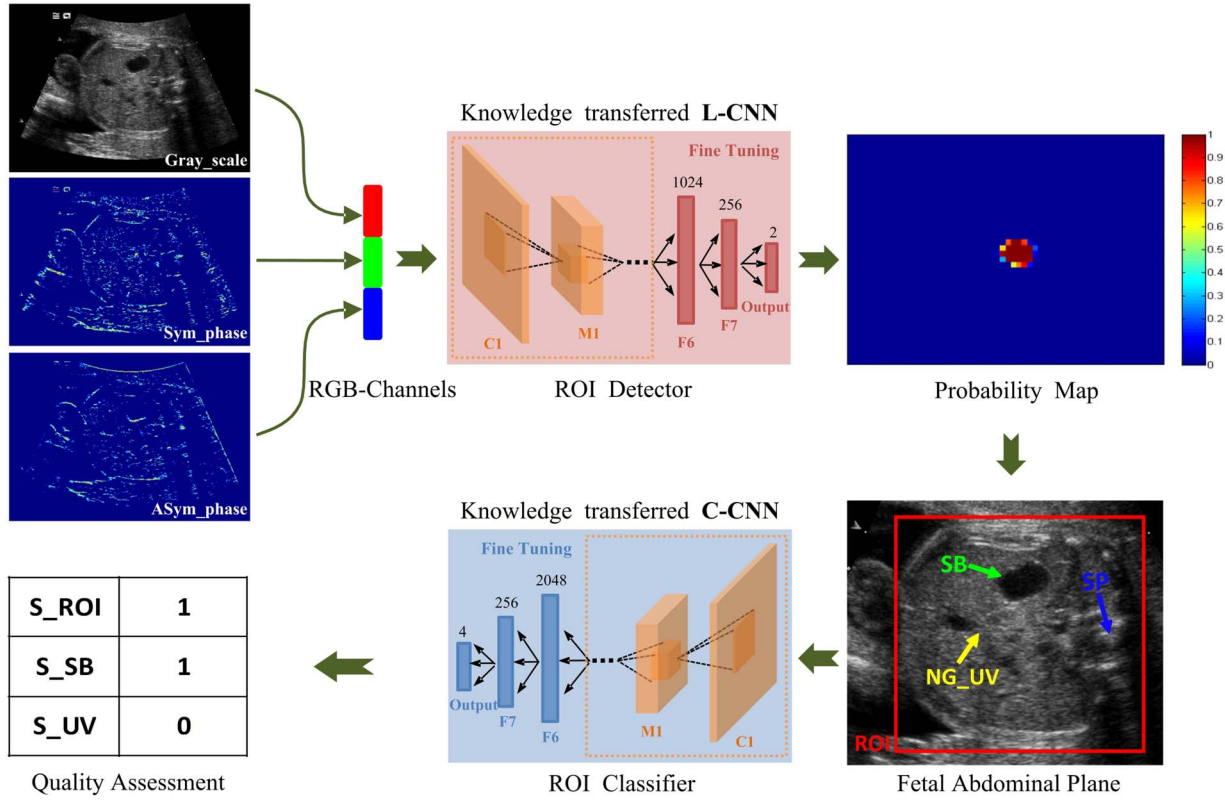
Fig. 3. Flowchart of the DCNN-based FUIQA scheme. For the L-CNN, the input sources include the original US image, symmetric and asymmetric phase features. The L-CNN will help to locate the ROI of fetal abdominal region, which is the input of the C-CNN. The knowledge learned from the L-CNN will be introduced to the C-CNN as initialization for the learning of four-class differentiation of US images. The C-CNN rates the input US image with the score of 2, since the UV is not well depicted.

TABLE II
ARCHITECTURE OF THE L-CNN MODEL

Layer	Feature maps	Kernel size	Stride
input	227x227x3	-	-
C1	55x55x96	11	4
M1	27x27x96	3	2
C2	27x27x256	5	1
M2	13x13x256	3	2
C3	13x13x384	3	1
C4	13x13x384	3	1
C5	13x13x256	3	1
M5	6x6x256	3	2
F6	1024	-	-
F7	256	-	-
F8	2	-	-

natural image domain. It has been shown that the low-level local cues learned from natural images are effectively transferrable to the domain of fetal US images for standard plane retrieval [10]. The setting of the L-CNN architecture is shown in Table II. Specifically, the *C* and *M* layers are initialized with the AlexNet whereas the three *F* layers are initialized with Gaussian randomizer. With the US training data, the L-CNN can be further fine-tined. For better generalization ability, the dropout strategy [34] and ReLu are implemented for training the L-CNN. The learning rate is set as 0.001 and will be gradually decreased by factor of 10 until the convergence is reached.

With the trained L-CNN, the ROI that encloses the fetal abdominal region in a testing fetal US image is then sought with the steps of center localization and rectangle definition. Center localization is realized with the sliding-window scanning of the testing US image and by computing the ROI probability of each window with the L-CNN. The sliding-window size is 320×280 and each cropped window is further resized to 227×227 to fit the L-CNN. Afterward, an ROI probability map can be derived (see Fig. 3). The global maximum of the ROI probability map is then regarded as the preliminary ROI center, denoted as $C_{x,y}$.

To further define the complete ROI, we perform second pass of window sliding within the restricted area centered at the $C_{x,y}$ with vertical and horizontal offsets of 10 pixels. We vary the size of sliding-window with width range of 210–400 and height range of 190–340 to find the proper size setting. The window with maximal probability from the L-CNN is then recommended as the final fetal abdominal ROI in the current testing image.

C. ROI Classification With the C-CNN

The C-CNN aims to assess the goodness of depiction for the SB and UV structures in the fetal US images. In clinical practice, the goodness of depiction for anatomical structures can be rated as "satisfactory," "NG," and "absent." The satisfactory depiction of structures suggests the reliability for the biometric measurements, whereas the NG structures may be short of complete presentation and may potentially lead to incorrect biometric measurements [1]. In this paper, an image
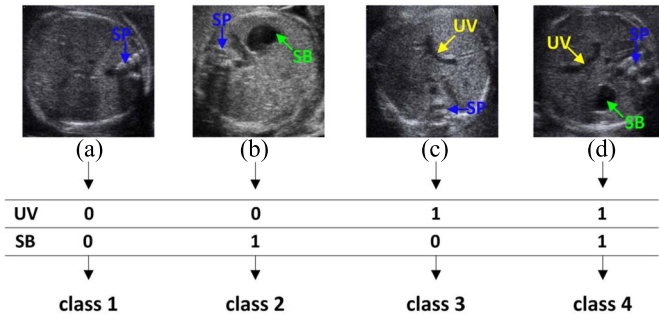
Fig. 4. Demonstration of four ROI classes. (a) Absent of SB and UV. Criterion of (b) SB satisfied and (c) UV satisfied. (d) Criteria of SB and UV satisfied. The cases of (a)–(d) are denoted as classes 1–4, respectively.

is given a score of 1 if either the C-CNN or the physicians rate it as satisfactory for criterion 2 or 3 in Table I. Other NG or absent images are scored as 0. Since we would like to evaluate the goodness of depiction of the SB and UV structures for the overall image quality assessment, the evaluation process is formulated as four-class classification problem for the C-CNN model. Fig. 4 lists the four exemplary statuses of goodness of depiction that stand for the four classes in the classification problem.

The reason of jointly casting the quality assessment with respect to the second and third criteria as the four-class classification scheme is that the SB and UV structures shall be co-presented in a qualified fetal US image. With the four-class learning framework, we expect the C-CNN architecture can automatically explore the spatial relation among the SB, UV, and SP structures and also tell the difference between the US planes with co-occurrences of SB, UV, and SP and the planes with co-present of SB, GB, and SP, as shown in Fig. 2(c). The SB-GB-SP planes are regarded as the samples of the C-CNN's second class as the UV structure is incorrect. Meanwhile, the US planes with all possible combinations of NG UV and NG SB structures are taken as the samples of the first, second, and third classes to teach the C-CNN. Since all three criteria in Table I are related to each other, the knowledge encoded in the L-CNN is transferred to the C-CNN to assist the four-class classification. Specifically, we allow the C-CNN share the same basic architecture of C and M layers with the L-CNN, but have different neural structures at the F layers for different reasoning purpose. The C-CNN has three F layers, which are constituted of 2048, 256, and 4 neurons, respectively. The network of C and M layers in the C-CNN is initialized with the settings of the constructed L-CNN, while the network of F layers is randomly set from Gaussian distribution. Meanwhile, the dropout strategy and dynamic adaptation of the learning rate of the L-CNN are also implemented for the C-CNN. The topmost F layer of the C-CNN is further connected with soft-max layer to yield the classification output. At the testing phase, we simply feed the ROI identified by the L-CNN to the C-CNN to obtain the classification result in terms of the second and third criteria in Table I.

D. Fetal Abdominal US Image Scoring

Referring to the Table I, the first criterion is to evaluate the area ratio of the depicted abdominal region. Therefore, the

score of the first criterion, S_{ROI}, can be determined by the ratio of the area of ROI to the area of field of view (FOV), denoted as $R_{r/FOV}$, with the differentiation rule defined in (4). The ROI from the L-CNN is then examined with the C-CNN and the corresponding scores for the second and third criteria, S_{SB} and S_{UV}, can be obtained as shown in Fig. 4. The overall quantitative assessment for an US fetal abdominal image, S_{FAP}, can be calculated with

$$S_{ROI} = 1, \text{ if } R_{r/FOV} \geq \frac{1}{2}; \text{ otherwise, } S_{ROI} = 0 \quad (4)$$

$$S_{FAP} = S_{ROI} + S_{SB} + S_{UV}. \quad (5)$$

E. Incorporation of Data Cropping and Flipping Procedure

In many DCNN implementations, a data cropping and flipping procedure [27] is commonly incorporated to artificially increase the size of the dataset. Specifically, the five patches with the size of 227×227 aligned at the four corners and the center of the original 256×256 ROI will be cropped as the augmented samples. Meanwhile, the same cropping operation is also applied on the flipped ROI along the horizontal direction to collect another five 227×227 patches. At the training of a DCNN, the data cropping and flipping are performed randomly, whereas the classification results of ten patches for a sample are averaged for the testing of a DCNN. In our experiments, it will be shown that the incorporation of the data cropping and flipping procedure can boost the classification performance of the C-CNN but turns out to be not helpful for the L-CNN. The underlying causes for the opposite performance influence will be analyzed in the discussion section.

F. Alternative Input Channels of DCNN Model With Local Phase Analysis

The DCNN model is originally devised for the analysis of natural images and has three input channels for the red, green, and blue components, respectively. Since the US images have simply single channel, one intuitive way to apply DCNN model on the analysis of US images is to duplicate the US images for the RGB channels. In this paper, we further explore alternative input channels with local phase features, which were shown to be effective for many US image processing problems [7], [30]–[32]. In particular, the phase symmetric feature was shown to be helpful for the detection of key abdominal structures in the fetal US images [7] and may be worth of exploration for our problem.

We specifically exploit the local phase symmetric and asymmetric features as the alternative input channels to see if the combination of original US content with these local phase features can boost the classification performance of the L-CNN and C-CNN. The phase symmetric feature at the pixel (x, y), denoted as $PS(x, y)$, can be computed as

$$PS(x, y) = \frac{\sum_r \sum_m \lfloor [|e_{r,m}(x, y)| - |o_{r,m}(x, y)|] - T_r \rfloor}{\sum_r \sum_m \sqrt{e_{r,m}(x, y)^2 + o_{r,m}(x, y)^2} + \varepsilon} \quad (6)$$

where $e_{r,m}(x, y)$ and $o_{r,m}(x, y)$ are the even- and odd-symmetric filter outputs at the scale m and orientation r, respectively. The term T_r is a noise compensation, whereas
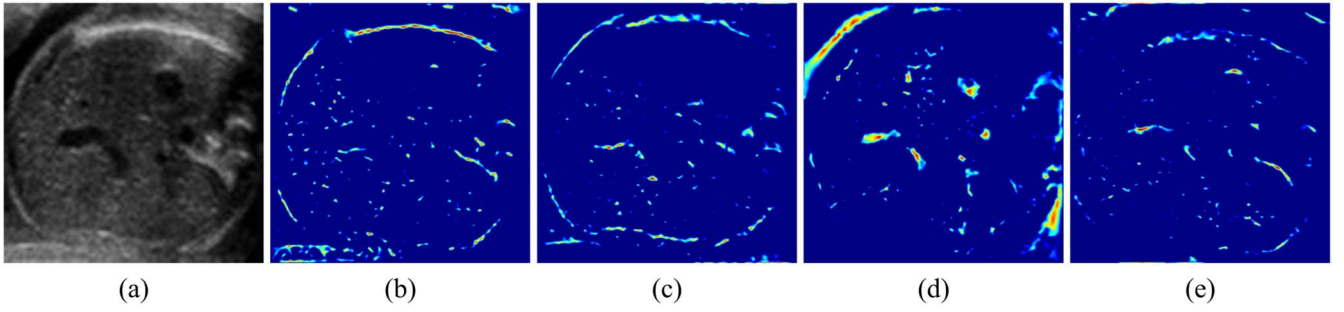
Fig. 5. Demonstration of local phase features. (a) Good fetal abdominal US image. Bright phase (b) symmetric map and (c) asymmetric map. Dark phase (d) symmetric map and (e) asymmetric map.

the term ε aims to prevent the situation of division by zero. The even- and odd-symmetric filters were suggested to be 2-D log-Gabor filters [31]. The phase symmetric feature defined in (6) can manifest either bright or dark symmetric components in the images. For only identification of bright symmetric components, it can be achieved by simply changing the numerator term $|e_{r,m}(x, y)| - |o_{r,m}(x, y)|$ as $e_{r,m}(x, y) - |o_{r,m}(x, y)|$, whereas the dark symmetric components can be computed with the replacement of numerator term as $-e_{r,m}(x, y) - |o_{r,m}(x, y)|$.

Similar to the definition of phase symmetric filtering, the phase asymmetric response for both bright and dark components at the pixel (x, y), denoted as $\text{PAS}(x, y)$, can be computed as

$$\text{PAS}(x, y) = \frac{\sum_r \sum_m \lfloor [|o_{r,m}(x, y)| - |e_{r,m}(x, y)|] - T_r \rfloor}{\sum_r \sum_m \sqrt{e_{r,m}(x, y)^2 + o_{r,m}(x, y)^2 + \varepsilon}}. \quad (7)$$

The odd-symmetric filter response is expected to be larger than the response of even-symmetric filter for the computation of phase asymmetric feature. The bright asymmetric components in the images can be manifested by substituting the numerator $|o_{r,m}(x, y)| - |e_{r,m}(x, y)|$ as $o_{r,m}(x, y) - |e_{r,m}(x, y)|$, while the replacement with $-o_{r,m}(x, y) - |e_{r,m}(x, y)|$ can obtain only the dark asymmetric components. More details about the phase symmetric and asymmetric feature computation can be found in [28] and [29].

Fig. 5 demonstrates the symmetric/asymmetric phase features with respect to the bright and dark components from a good fetal abdominal US image. As can be observed, bright local phase features can manifest the AC border and may be helpful for the task of ROI localization. The dark local phase features aim to enhance image cues related to the dark structures of SB and UV to assist the implementation of criteria 2 and 3 in Table I. We will perform a comprehensive study on several channel input replacement options with the phase symmetric/asymmetric features on the architectures of the L-CNN and C-CNN in the experiments and give the related discussion in the discussion section.

III. EXPERIMENTS AND RESULTS

A. Image Data for Training and Testing

All the fetal US abdominal images in this paper were acquired from the Shenzhen Maternal and Child Healthcare

TABLE III
NUMBER DISTRIBUTIONS OF TRAINING AND VALIDATION GROUPS FOR THE TRAINING OF THE L-CNN AND C-CNN

Group	L-CNN		C-CNN			
	Positive	Negative	Class 1	Class 2	Class 3	Class 4
Training	13968	14695	3358	3472	3541	3597
Validation	1447	1855	359	355	366	367

Hospital during September 2012 to November 2013. The US images were recorded with conventional hand-held 2-D US probe on pregnant women in the supine position, by following the standard obstetric examination protocol. All US images were acquired with a Siemens Acuson Sequoia 512 US scanner. The fetal gestational ages of all subjects range from 16 to 40 weeks. All data were acquired with various settings of imaging parameters, e.g., time gain compensation, for the best convenience of examination. In this paper, 492 US videos collected at the year of 2012 are treated as the source of training data, whereas 219 videos acquired in 2013 are taken as the testing data source.

8072 fetal abdominal images from 492 US videos are involved for the training of the L-CNN model to locate the ROIs of fetal abdomen in the US images. One positive ROI sample is extracted from each image of the training data. The negative ROI samples are randomly collected from the background of the images. A negative ROI sample may overlap with the positive sample no more than 40% to consider the case of partial coverage of fetal abdominal region. All positive training samples of the L-CNN are further treated as the overall training samples of the C-CNN. There are totally 1256, 3827, 151, and 2838 samples with respect to the classes 1–4. In order to balance the data among the four classes, we augment the data size of minority classes, i.e., classes 1, 3, and 4, by randomly rotating the class samples with the degree range of (-25, 25) as new samples. Afterward, there are totally 15 415 positive ROI samples for the training of the L-CNN model, including 3717, 3827, 3907, and 3964 samples with respect to the classes 1–4 for the training of the C-CNN model. The annotation of all training data for the L-CNN and C-CNN was initially done by a graduate student with two years of experience on fetal US images. The annotation data were further reviewed and revised by our clinical advisor committee to ensure the correctness. In the training of the L-CNN and C-CNN, we further split the original training data into the
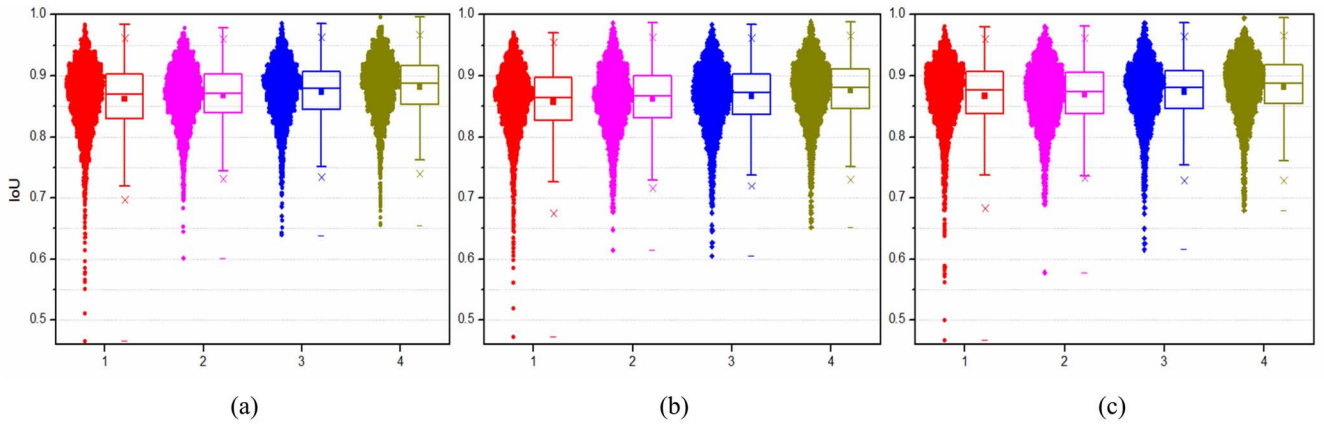
Fig. 6. Box-plots of IoU values to compare the ROI localization results from four L-CNN implementations with manually defined ROI sets from medical doctors of $E1$–$E3$. Comparison to (a) $E1$, (b) $E2$, and (c) $E3$. In each subfigure, the IoU distribution of every experiment is put on the left hand side of each box-plot. The distributions and box-plots with colors of red, magenta, blue, and green are drawn from the implementations of experiment 1 (crop), experiment 2 (no_crop), experiment 3 (no_crop_sym_bright), and experiment 4 (no_crop_sym_asym_bright), respectively.

training and validation groups with the approximate ratio of 9 to 1 (see Table III for details).

The testing data for the L-CNN and C-CNN models include 2606 fetal abdominal images from 219 US videos acquired from 66 women at pregnancy. Each testing image is rated by three medical doctors with experience on fetus US examination more than three years by following the quality control criteria defined in Table I. The three medical doctors are denoted as $E1$–$E3$ throughout this paper. Specifically, three sets of manually defined fetal abdominal ROIs in the 2606 images are extracted by the three medical doctors for the evaluation of the L-CNN, whereas the goodness of depiction with respect to the UV and SB structures is assessed by the three medical doctors as the evaluation references for the C-CNN model. All ROI annotations and UV/SB rating processes of each medical doctor were conducted independently.

B. System Implementation and Evaluation

We employ the Caffe implementation of the DCNN architecture [35] for the L-CNN and C-CNN models. At the testing stage, it generally takes around 1 min for the ROI localization, evaluation of the two anatomical structures and fetal abdominal image scoring in a workstation equipped with 2.60 GHz Intel Xeon E5-2670 CPU and a NVIDIA GF100GL Quadro 4000 GPU.

To demonstrate the efficacy of the FUIQA scheme of fetal abdominal US images, the intermediate and final scoring results are compared to the corresponding manual annotations and ratings from the three radiologists [36]. The experimental evaluations of our FUIQA scheme are carried out in three parts. The first part of evaluation is to investigate the two implementation issues for the L-CNN and C-CNN. Second, we qualitatively illustrate the performance of the C-CNN model with visualization of the model discriminative capability before and after training. Meanwhile, several results from our FUIQA scheme are demonstrated for visual observation on the performance of image quality assessment. In the third part,

we evaluate each technical component and the final scoring performance with quantitative metrics.

1) Implementation Issues for the L-CNN and C-CNN: In this section, we explore which implementation option can be most beneficial for the L-CNN and C-CNN. For L-CNN, we evaluate four implementation options with four corresponding experiments, denoted as "crop," "no_crop," "no_crop_sym_bright," and "no_crop_sym_asym_bright," respectively. Specifically, the crop experiment implements the random data cropping procedure that automatically crops 227×227 patches centered at the center and four corners of the original and mirrored input 256×256 images. The data cropping procedure of the deep learning model aims to augment the training data. The no_crop experiment does not implement the data cropping procedure but only includes the original images by resizing them into 227×227. The no_crop_sym_bright experiment also does not implement the data cropping procedure but replaces the second input channel of the L-CNN architecture with the bright phase symmetric map. Since the definition of fetal abdominal ROI can be referred to the circumference borders, which appear as bright cues in the US images, we only investigate on bright symmetric and asymmetric features for the L-CNN. Based on the setting of no_crop_sym_bright experiment, the no_crop_sym_asym_bright experiment replaces the third input channel with the bright phase asymmetric map, and therefore the three input channels carry heterogeneous image contents.

The performance of the four experiments is evaluated with the metrics of intersection over union (IoU) to illustrate the goodness of ROI localization with the four variational implementation options. Given a computerized ROI A and a manually defined ROI B, the IoU metric can be computed as the ratio of the area of intersection to the area of union, like

$$IoU = (A \cap B)/(A \cup B). \qquad (8)$$

Fig. 6 illustrates the box-plots of the IoU values for the comparison of the identified ROIs by the L-CNN implementations and the manual ROIs defined by $E1$–$E3$, respectively.
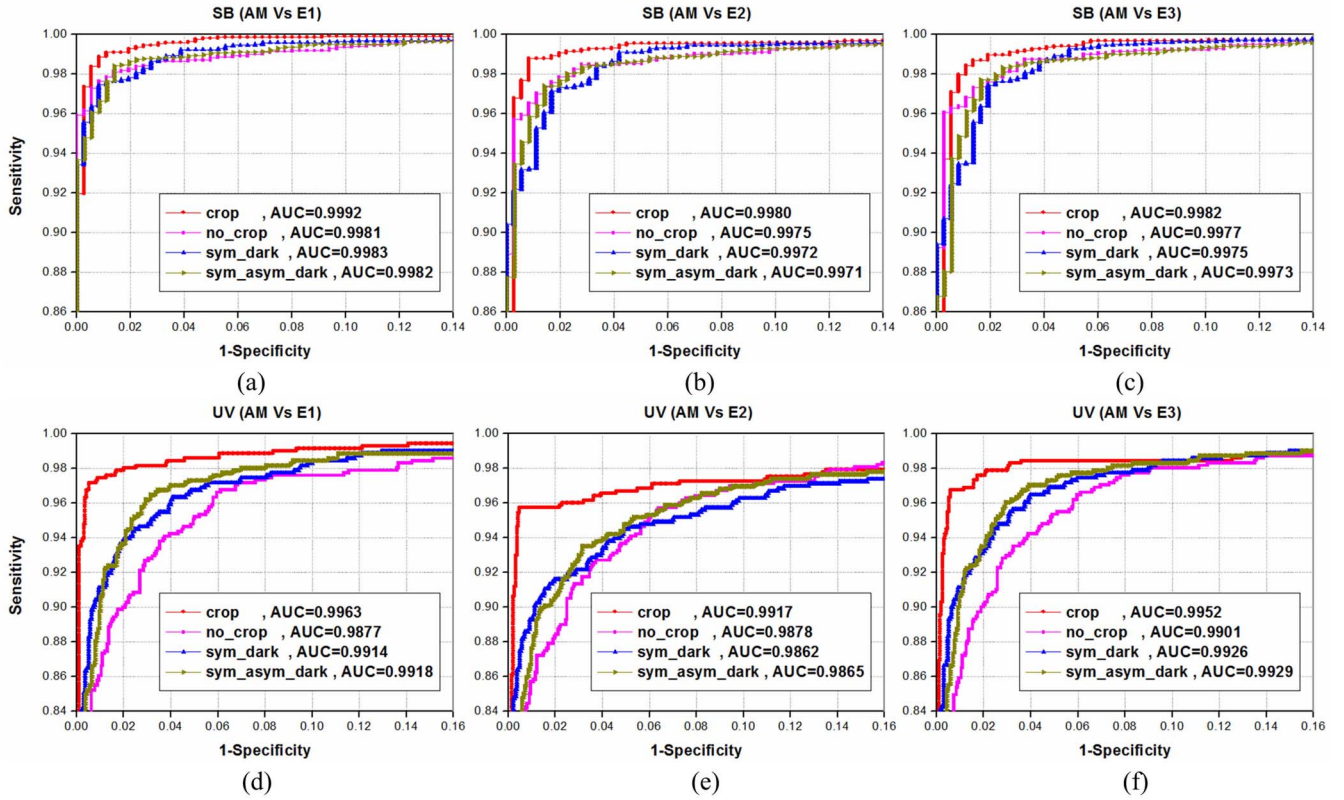
Fig. 7. AUC values and ROC curves for the comparison of four implementations of C-CNN. (a)–(c) SB performance comparison to the medical doctors of $E1$–$E3$, respectively. (d)–(f) UV performance comparison to the medical doctors of $E1$–$E3$, respectively. In all subfigures, curves colored in red, magenta, blue and green stand for the performances of the experiment 1 (crop), experiment 2 (no_crop), experiment 3 (sym_dark), and experiment 4 (sym_asym_dark), respectively.

Similar to the L-CNN, we evaluate four implementation options of the C-CNN with four corresponding experiments: 1) "crop"; 2) "no_crop"; 3) "sym_dark"; and 4) "sym_asym_dark." The variational implementations are mostly like the four implementations of the L-CNN. One difference is that we focus on dark structures but not bright components in the images. Since the target structures, i.e., UV and SB, appear as dark objects in US images, we only compute the local phase features for dark components as the alternative input sources for the C-CNN implementations. On the other hand, it will be shown the data augmentation does help for the C-CNN, and therefore the data augmentation procedure is implemented in the incorporation options of local phase features, i.e., sym_dark and sym_asym_dark. Meanwhile, as the C-CNN model is designed for the four-class classification, we adopt the area under ROC curve (AUC) to evaluate the performance of four C-CNN implementations. The performance comparison for the four experiments of the C-CNN with respect to UV and SB structures is shown in Fig. 7 with AUC values and ROC curves.

2) Qualitative Performance Analysis: To qualitatively illustrate the effectiveness of the learned features from the C-CNN, the *t*-distributed stochastic neighbor embedding (*t*-SNE) technique [37] is employed for visual inspection. The *t*-SNE technique is a dimensionality reduction method which is able to embed the high-dimensional feature space into 2-D or 3-D for the visualization of discriminative capability. Specifically,

we use the *t*-SNE to illustrate the feature discriminative capability of the *n*F7 layer as well as the raw input image layer of the C-CNN architecture with the training and testing data. Fig. 8 depicts the distributions of the four classes of the C-CNN model in the 3-D embedded visualization space. The data samples of classes 1–4 are presented in red, blue, green, and cyan colors, respectively, in Fig. 8.

3) Quantitative Performance Evaluation: To quantitatively illustrate the performance of the L-CNN and C-CNN, we compare the results of ROI identification and the assessment of SB and UV structures from our FUIQA scheme with the annotations and ratings from three radiologists. Three assessment metrics of accuracy, sensitivity and specificity are adopted for quantitative comparison. Table IV reports the performance comparison between the results from our FUIQA scheme with each set of expert's annotations using the three assessment metrics in terms of ROI, SB and UV criteria. The Cohen's kappa [38] is also employed here to evaluate the agreement not only between our FUIQA results and the radiologists' quality assessment results but also the agreement among the radiologists' ratings. The corresponding Cohen's kappa statistics can be found in Table V.

To illustrate the single efficacy of the C-CNN, the C-CNN is applied on the manual ROIs instead of the ROIs generated by the L-CNN for comparison. Table VI shows the performance of the C-CNN with the inputs of manual ROIs. Since the C-CNN exploits the interaction between the

TABLE IV
PERFORMANCE COMPARISON BETWEEN THE FUIQA RESULTS AND ANNOTATIONS FROM $E1$–$E3$. THE ABBREVIATIONS OF
"ACC," "SEN," AND "SPEC" STAND FOR ACCURACY, SENSITIVITY, AND SPECIFICITY

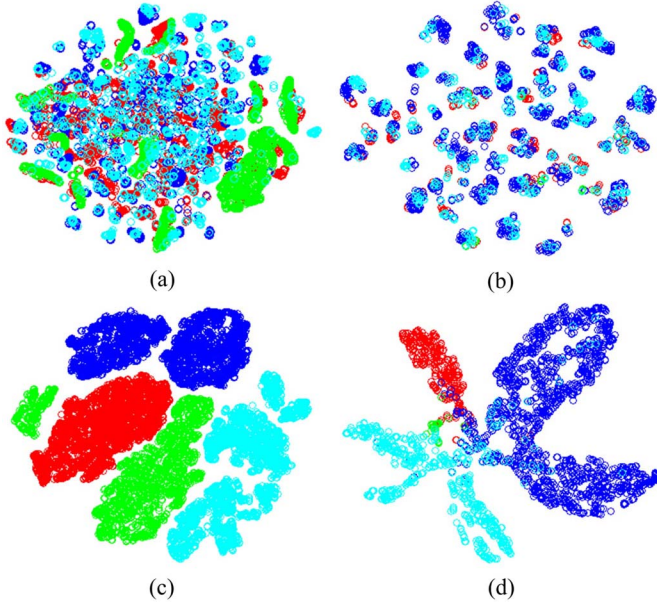|	ROI			SB			UV		
	Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec
FUIQA/E1	0.930	0.949	0.878	0.990	0.992	0.975	0.983	0.958	0.992
FUIQA/E2	0.915	0.954	0.819	0.987	0.990	0.967	0.976	0.935	0.992
FUIQA/E3	0.920	0.968	0.812	0.986	0.991	0.956	0.980	0.952	0.991

Fig. 8. Visualization with the t-SNE technique for samples of four classes. The samples colored in red, blue, green, and cyan are of classes 1–4, respectively. Distribution of (a) training samples in the embedded 3-D space with raw image features, (b) testing samples in the embedded 3-D space with raw image features, (c) training samples in the embedded 3-D space with nF7 layer features, and (d) testing samples in the embedded 3-D space with nF7 layer features.

TABLE V
AGREEMENT AND COHEN'S KAPPA VALUES BETWEEN THE RESULTS
FROM OUR FUIQA SCHEME AND MEDICAL DOCTORS ($E1$–$E3$)

	FUIQA/E1	FUIQA/E2	FUIQA/E3	E1/E2	E1/E3	E2/E3
Agreement	0.907	0.882	0.892	0.929	0.924	0.929
Kappa	0.851	0.813	0.830	0.888	0.881	0.888

TABLE VI
PERFORMANCE OF THE C-CNN WITH THE INPUTS OF
MANUALLY LOCATED ROIs FROM $E1$–$E3$

	SB			UV		
	Acc	Sen	Spec	Acc	Sen	Spec
Manual_ROI/E1	0.989	0.992	0.967	0.983	0.969	0.987
Manual_ROI/E2	0.987	0.991	0.963	0.973	0.949	0.980
Manual_ROI/E3	0.988	0.992	0.964	0.981	0.969	0.986

TABLE VII
PERFORMANCE COMPARISON BETWEEN THE FUIQA RESULTS AND
ANNOTATIONS FROM $E1$–$E3$ USING SEPARATE
CLASSIFICATION SCHEME

	SB			UV		
	Acc	Sen	Spec	Acc	Sen	Spec
Separate_classify/E1	0.985	0.991	0.942	0.965	0.927	0.979
Separate_classify/E2	0.983	0.990	0.941	0.960	0.908	0.980
Separate_classify/E3	0.984	0.992	0.934	0.961	0.920	0.977

TABLE VIII
AGREEMENT AND COHEN'S KAPPA VALUES BETWEEN THE RESULTS
FROM OUR FUIQA SCHEME AND MEDICAL DOCTORS USING
SEPARATE CLASSIFICATION SCHEME

	Separate_classify/E1	Separate_classify/E2	Separate_classify/E3
Agreement	0.882	0.861	0.871
Kappa	0.812	0.780	0.797

TABLE IX
PERFORMANCE COMPARISON BETWEEN THE FUIQA RESULTS AND
ANNOTATIONS FROM $E1$–$E3$ USING THE NEW CRITERION
FOR SCORING A COMPUTERIZED ROI

	ROI		
	Acc	Sen	Spec
FUIQA/E1	0.928	0.946	0.878
FUIQA/E2	0.913	0.951	0.819
FUIQA/E3	0.918	0.965	0.812

TABLE X
AGREEMENT AND COHEN'S KAPPA VALUES BETWEEN THE RESULTS
FROM OUR FUIQA SCHEME AND MEDICAL DOCTORS USING THE
NEW CRITERION FOR SCORING A COMPUTERIZED ROI

	FUIQA/E1	FUIQA/E2	FUIQA/E3
Agreement	0.905	0.880	0.891
Kappa	0.849	0.812	0.828

UV and SB structures implicitly in the four-class classification scheme, we further implement a separated scheme that trains the classifiers of the UV and SB structures separately to show if the joint consideration on these two structures using the four-class classification scheme can be effective. The performance of the separated classification scheme can be found in Tables VII and VIII. In the final experiment, we explore the issue of the correctness of fetal ROI on the first ROI criterion defined in Table I. Specifically, we implement a new rigorous ROI criterion by considering not only the area ratio of (4) but also the correctness of computerized ROIs. A correct computerized ROI is supposed to cover the entire fetal abdominal region. The determination of the ROI correctness is carried by an expert with two years of experience in fetal US. The evaluation results of the correctness by the expert are further incorporated into the new ROI criterion for the criterion 1 scoring. Tables IX and X report the relevant performance of the new rigorous criterion 1 implementation.
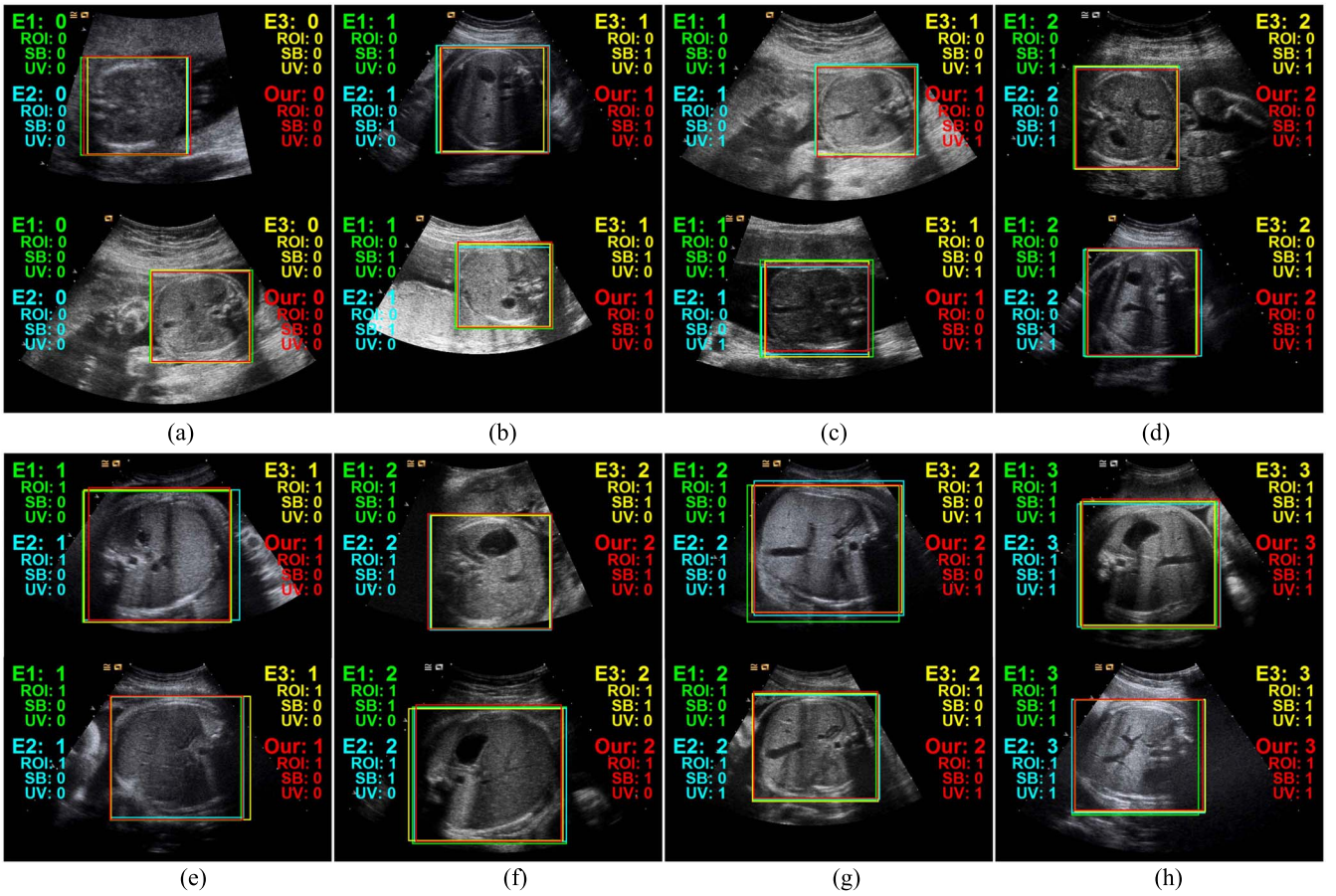
Fig. 9. Demonstration of our results that perfectly match with the annotations from *E*1–*E*3. The ratings and ROIs are colored in green, blue, yellow, and red are from *E*1–*E*3 and our method (our for short), respectively.

IV. DISCUSSION

Referring to Fig. 6, the goodness of the identified ROIs by the L-CNN can be gradually improved with every change of implementation from the experiments 1–4. For instance, the average values between our method and *E*1 for the four implementations are 0.863, 0.868, 0.873, and 0.882, respectively, suggesting the best performance can be achieved with the experiment 4. Therefore, the best configuration for the L-CNN implementation is no_crop_sym_asym_bright. The incorporation of data augmentation degrades the performance of fetal abdominal region localization. It may be because the sampled data aligned at the four corners of the original ROI may not always cover the entire fetal abdominal region. Partial AC borders may be missed. Therefore, the L-CNN will be misled by these NG augmented data and produce worse performance.

For the implementation of the C-CNN, the ROC curves shown in Fig. 7 suggest that the best assessment performance for either UV or SB can be achieved with the implementation of crop. It may be because the data augmentation procedure can enrich the training data with diversity and then endows the C-CNN more generalization capability. The targeted structures of UV, SB, and SP in the data for training the C-CNN are away from the circumference border and have less chance to be partially cropped away in the augmented data. Therefore, the data augmentation appears to be helpful. On the other hand, the

AUC differences of SB structure with respect to each medical doctor are around 0.001 whereas the differences of UV structure are around 0.005. It may suggest that the critical structure that leads to performance differences for the four C-CNN implementations is UV.

In Fig. 8, either the training or testing samples of the four classes with the raw input layer are mixed with each other. It thus suggests the intraclass variation is quite high and interclass variation is pretty low for the four-class classification in terms of the raw US image features. In contrast, the training and testing samples of the four classes appear to be better separated with the features of *n*F7 layer in the C-CNN. Accordingly, with the basis of the learned *n*F7 layer in the C-CNN architecture, better classification results may be obtained more easily.

Fig. 9 demonstrates several typical cases on which our FUIQA scheme shares the same rating agreement with the three medical doctors. In Fig. 9(a)–(d), the areas of the depicted abdominal ROIs are smaller than 1/2 of the US FOV and thus these cases are rated as 0 by our system for the first criterion of Table I. On the other hand, the abdominal ROI areas in Fig. 9(e)–(h) are identified as larger than 1/2 of the US FOV and hence these cases are scored as 1. The SB structures in Fig. 9(b) and (f) are scored as 1 with respect to the second criterion. However, the UV structures are regarded
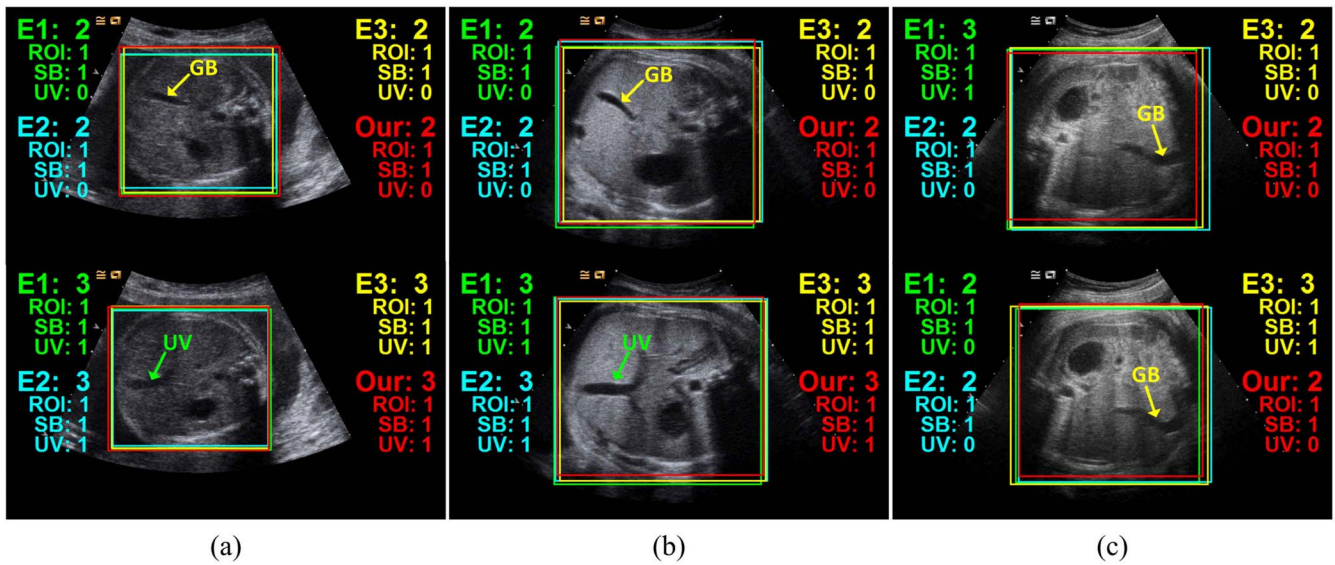
Fig. 10.   Cases that our system can correctly differentiate UV and GB structures.
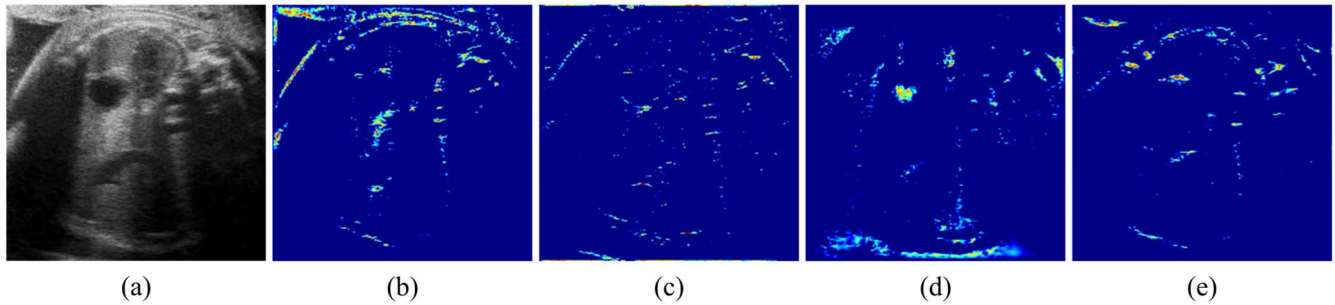

Fig. 11.   Illustration of an US image with low quality and its local phase feature maps. (a) US image. Bright phase (b) symmetric map and (c) asymmetric map. Dark phase (d) symmetric map and (e) asymmetric map.

as NG by our C-CNN for the cases in Fig. 9(b) and (f). The UV structures in Fig. 9(c) and (g) are scored 1 by our C-CNN, because the shapes and margins satisfy the third criterion. Unfortunately, the C-CNN regards the SB structures in Fig. 9(c) and (g) as NG and hence these two cases are eventually classified as class 3 by our system. Both SB and UV are satisfactorily depicted in the cases of Fig. 9(d) and (h), whereas the ROI criterion is met for the case in Fig. 9(h) but not the case in Fig. 9(d). Accordingly, the cases of Fig. 9(h) can be recommended for the biometric measurement. It is worth noting that the C-CNN results shown in Figs. 8 and 9 are from the implementation of crop. Meanwhile, as shown in Fig. 9, our DCNN-based FUIQA scheme is robust to various imaging factors like time gain compensation and fetal pose. Furthermore, the FUIQA may also achieve correct quality assessment on images with serious acoustic shadow [see Fig. 9(b), (d), and (h) for examples]. The DCNN model can implicitly consider the variation of imaging parameter setting, fetal position and shadowing effect from the diverse training data. With conventional image processing and pattern recognition methods, these factors shall be explicitly and carefully tackled.

Since the realization of criteria 2 and 3 is within the four-class classification framework, it allows the C-CNN to automatically explore the relative geometry relation of the SP, SB and UV structures. The reasoning about the relative relation among fetal structures can equip the C-CNN with the capability for the differentiating similar structures of UV and GB (see Fig. 10). The upper and lower cases in Fig. 10(a) and (b) were acquired from the same subject at the same arrival of examination but with different radiologists. As can be found, the scores yielded from the C-CNN match with the scores from the three medical doctors, even for the low contrast cases shown in Fig. 10(a). On the other hand, the C-CNN can still achieve correct scoring for the difficult cases shown in Fig. 10(c), where the medical doctors $E1$ and $E3$ mistakenly identify GB as UV. In this case, our automatic FUIQA scheme may help to compensate the human error.

Fig. 11 illustrates a challenging low-quality case of fetal abdominal US image along with the computed local phase features. As can be found in Fig. 11(a)–(c), albeit serious shadowing effect, some AC border cues can still be captured. On the other hand, the UV structure is not able to be identified with phase symmetric feature [see Fig. 11(d)] due to the low contrast of image quality and shadowing effect. Accordingly, compared to the UV structure, the cues of circumference border are relatively easier to be manifested by local phase features. The local phase features are shown to be helpful for
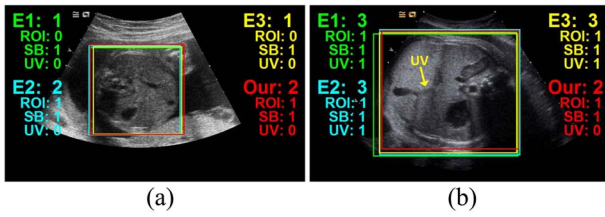
Fig. 12. Difficult cases that our system cannot reach the same scoring with the three medical doctors.

the L-CNN, as the bright phase symmetric and asymmetric features are able to manifest importance cues like the circumference border (see Fig. 5) and exclude redundant image cues. For the C-CNN, we compute the dark phase symmetric and asymmetric features to character the targeted structures of UV and SB. However, the local phase features turn out to slightly lower down the C-CNN performance. Referring to Figs. 5 and 11, unlike the SB structure, the UV structure cannot be stably identified by dark phase symmetric and asymmetric features. Therefore, the performance of UV structure assessment is lightly degraded, as the dark phase features may confuse the C-CNN. The original US images, on the other hand, are sufficient to support the four-class differentiation with promising performance.

Fig. 12 demonstrates several difficult cases which our FUIQA scheme cannot yield consistent scores with the ratings of $E1$–$E3$. In Fig. 12(a), it can be observed that the ROI derived from our FUIQA scheme is quite close to the ROIs defined by $E1$–$E3$, but the corresponding computerized score does not match with the experts' ratings. The $R_{r/\text{FOV}}$ ratios from $E1$–$E3$ and our FUIQA scheme for the case of Fig. 12(a) are 0.497, 0.513, 0.476, and 0.523, respectively. Accordingly, the inconsistent scoring between the experts and our system could also possibly stem from the hard thresholding [see (4)]. A shadow-degraded UV structure is presented in Fig. 12(b). The middle part of the UV is blurred by the acoustic shadow but all medical doctors are able to recognize the curve shape of the UV structure. Our system seems to be unable to identify the upper portion of this UV structure and therefore reaches inconsistent score with the experts' ratings.

Referring to Tables IV and VI, it can be found that the performances from the computerized and manual ROIs are very close. It may suggest that the L-CNN can correctly localize the fetal abdominal ROI and meanwhile the C-CNN is robust to the variations of ROIs defined either by experts or the L-CNN. The efficacy of the joint consideration of UV and SB structures in the four-class classification scheme of the C-CNN can be observed by comparing the performances in Tables IV, V, VII, and VIII. The four-class classification scheme can achieve better performance than the separated scheme on the UV and SB structures. The performance of the new rigorous ROI criterion implementation appears to be quite close to the performance of the old ROI criterion (see Tables IV, V, IX, and X). It may thus suggest that most computerized ROIs scored in the ROI criterion are correct.

Fig. 13 lists the feature patterns of the L-CNN, whereas the patterns of the C-CNN can be found in Fig. 14. The feature
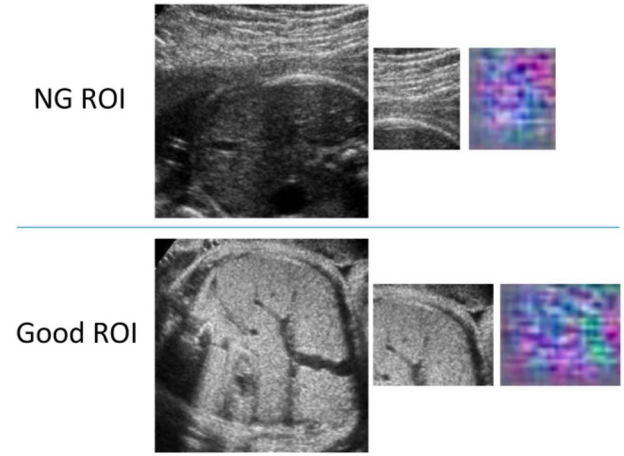

Fig. 13. Examples of learned patterns of the L-CNN for the classes of NG and good ROI are shown in the first and second rows.
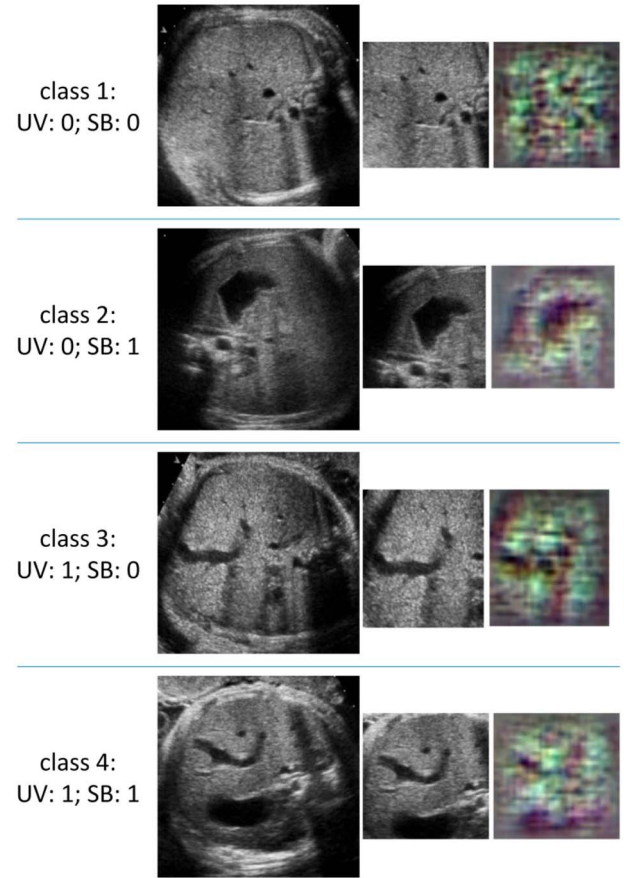

Fig. 14. Examples of learned patterns of the C-CNN for the four classes are shown in the four corresponding rows.

patterns are computed with the visualization technique [39]. For each row in Figs. 13 and 14, the input ROI is listed in the leftmost column, whereas the middle column is the sub-ROI that can produce the highest activated neuron in convolutional layer 5 that is shown in the rightmost column. The regions with purple color in the rightmost column suggest higher activation in the layer. As can be found in Fig. 13, the feature patterns characterize the nonfetal abdominal regions in NG

ROIs, whereas the feature patterns of good ROI try to emphasize the AC borders of fetus. On the other hand, since both UV and SB are absent or poorly defined in class 1 images, the feature patterns characterize the non-UV and non-SB portions. The feature patterns of the second and third classes aim to manifest the structures of SB and UV, respectively. On the other hand, the fourth class tries to emphasize both SB and UV structures with higher activations in the feature patterns.

The major computational bottleneck lies in the ROI localization. It takes multiple passes of sliding-window scanning to search the proper ROI that encloses the entire fetal abdominal region in an US image. Every sliding window is evaluated with the L-CNN model to determine whether the window is the desirable ROI. It approximately calls 3339 times of L-CNN evaluation with around 58.6 s for an image in a workstation equipped with 2.60-GHz Intel Xeon E5-2670 CPU and a NVIDIA GF100GL Quadro 4000 GPU. About 95% of the computation time of the FUIQA scheme is consumed by the ROI localization alone. The CPU performs relatively little computation, like data flow control, memory movement, GPU resource arrangement and so on. The future endeavor will be focused on the acceleration of ROI localization step with the parallelization programming skill or better GPU hardware equipment.

V. Conclusion

The quality control for the acquired fetal US images is crucial for the fetal biometrics measurements and prenatal diagnosis. However, human evaluation on the quality of fetus US images is subjective and highly depends on the evaluator's experience. Meanwhile, since the human evaluation involves the tasks of manual calculation of ROI area against the whole region of US FOV, the evaluation process can be very tedious and unfeasible in clinical practice. To improve the efficiency of quality control and offer objective assessment, an automatic FUIQA scheme based on DCNN model is developed in this paper. The developed FUIQA scheme may help to improve the fetal US acquisition outcomes of both novices and technicians. The proposed scheme may also be served as an assistive computerized toolkit for the education of radiology residents and medical students to sharpen their skill of US scanning for the fetus examination.

The FUIQA scheme has been evaluated with extensive experiments and shown to be comparable to manual assessments by experts. The FUIQA scheme is free of manual calculation of the ROI area ratio defined in the criterion 1 and the corresponding performance of the L-CNN is quite close to manual definitions. In terms of technical advance on the deep learning research on medical image analysis, a comprehensive exploration of alternative input channel sources with local phase features is conducted on the L-CNN and C-CNN models. Our experimental results suggest that the local phase features are helpful to improve the performance of the L-CNN but not the C-CNN. The proposed DCNN-based FUIQA can be easily generalized to other types of fetal images without the need of extra feature engineering.

References

[1] N. J. Dudley and E. Chapman, "The importance of quality management in fetal measurement," *Ultrasound Obstet. Gynecol.*, vol. 19, no. 2, pp. 190–196, 2002.

[2] J. E. Lawn, S. Cousens, and J. Zupan, "4 million neonatal deaths: When? where? why?" *Lancet*, vol. 365, no. 9462, pp. 891–900, 2005.

[3] G. C. S. Smith, M. F. S. Smith, M. B. McNay, and J. E. E. Fleming, "The relation between fetal abdominal circumference and birthweight: Findings in 3512 pregnancies," *BJOG Int. J. Obstet. Gynaecol.*, vol. 104, no. 2, pp. 186–190, 1997.

[4] Y. Ville, "'Ceci n'est pas une échographie': A plea for quality assessment in prenatal ultrasound," *Ultrasound Obstet. Gynecol.*, vol. 31, no. 1, pp. 1–5, 2008.

[5] L. Salomon *et al.*, "Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester," *Ultrasound Obstet. Gynecol.*, vol. 27, no. 1, pp. 34–40, 2006.

[6] B. Rahmatullah, I. Sarris, A. Papageorghiou, and J. A. Noble, "Quality control of fetal ultrasound images: Detection of abdomen anatomical landmarks using AdaBoost," in *Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro*, Chicago, IL, USA, 2011, pp. 6–9.

[7] B. Rahmatullah, A. T. Papageorghiou, and J. A. Noble, "Integration of local and global features for anatomical object detection in ultrasound," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Nice, France, 2012, pp. 402–409.

[8] D. Ni *et al.*, "Standard plane localization in ultrasound by radial component model and selective search," *Ultrasound Med. Biol.*, vol. 40, no. 11, pp. 2728–2742, 2014.

[9] H. Chen *et al.*, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.

[10] H. Chen *et al.*, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Munich, Germany, 2015, pp. 507–514.

[11] M. Wang *et al.*, "Facilitating image search with a scalable and compact semantic mapping," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1561–1574, Aug. 2015.

[12] S. Chen *et al.*, "Bridging computational features toward multiple semantic features with multi-task regression: A study of CT pulmonary nodules," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2016, pp. 53–60.

[13] S. Chen *et al.*, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images," *IEEE Trans. Med. Imag.*, to be published, doi: 10.1109/TMI.2016.2629462.

[14] B. Du *et al.*, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2536638.

[15] Y. Wei *et al.*, "Cross-modal retrieval with CNN visual features: A new baseline," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 1–12, Feb. 2017.

[16] J.-Z. Cheng *et al.*, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans," *Sci. Rep.*, vol. 6, Apr. 2016, Art. no. 24454.

[17] J. Shi *et al.*, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, Jun. 2016.

[18] F. Ciompi *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, 2015.

[19] H. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2015.

[20] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[21] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.

[22] Y. Guo *et al.*, "Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Boston, MA, USA, 2014, pp. 308–315.

[23] H. Chen, X. Qi, J.-Z. Cheng, and P.-A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Proc. 30th AAAI Conf. Art. Intell.*, Phoenix, AZ, USA, 2016, pp. 1167–1173.

[24] Y. Song *et al.*, "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 288–300, Jan. 2017.

[25] Y. Song *et al.*, "Segmenting overlapping cervical cell in pap smear images," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Prague, Czech Republic, 2016, pp. 1159–1162.

[26] X. Zhen *et al.*, "Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation," *Med. Image Anal.*, vol. 30, pp. 120–129, May 2016.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[28] P. Kovesi *et al.*, "Symmetry and asymmetry from local phase," in *Proc. 10th Aust. Joint Conf. Art. Intell.*, vol. 190. 1997, pp. 2–4.

[29] P. Kovesi, "Image features from phase congruency," *Videre J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.

[30] V. Grau and J. A. Noble, "Adaptive multiscale ultrasound compounding using phase information," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Palm Springs, CA, USA, 2005, pp. 589–596.

[31] I. Hacihaliloglu, R. Abugharbieh, A. J. Hodgson, and R. N. Rohling, "Bone surface localization in ultrasound using image phase-based features," *Ultrasound Med. Biol.*, vol. 35, no. 9, pp. 1475–1487, 2009.

[32] A. Belaid, D. Boukerroui, Y. Maingourd, and J.-F. Lerallut, "Phase-based level set segmentation of ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 138–147, Jan. 2011.

[33] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *Aistats*, vol. 15, no. 106, pp. 315–323, 2011.

[34] S. Wager, S. Wang, and P. S. Liang, "Dropout training as adaptive regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 351–359.

[35] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Orlando, FL, USA, 2014, pp. 675–678.

[36] S. Rueda *et al.*, "Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 797–813, Apr. 2014.

[37] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[38] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[39] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

Authors' photographs and biographies not available at the time of publication.