Automatic Identification of Multiple Planes of a Fetal Organ from 2D Ultrasound Images

Pradeeba Sridar[1], Ashnil Kumar[2], Ann Quinton[3,5], R. Krishna Kumar[1], Dagan Feng[2,4], Ralph Nanan[5], and Jinman Kim[2]

[1] Department of Engineering Design, Indian Institute of Technology Madras, India
[2] School of Information Technologies, The University of Sydney, Australia
[3] School of Medical and Applied Science, Central Queensland University, Australia
[4] Med-X Research Institute, Shanghai Jiao Tong University, China
[5] Charles Perkins Centre Nepean, The University of Sydney, Australia

Abstract. Retrospective analyses of ultrasound images (US) are required to correlate the prenatal growth parameters with health outcomes later in life. The reliability of such analyses depends upon the accurate identification of US image planes. However, the different planes of any single organ exhibit a high degree of intra-plane similarity in fetal US images. In this paper, we present a general framework to automatically identify the different planes of any single fetal organ. We fine-tune a pre-trained convolutional neural network (CNN) to create a feature extractor that derives the image features that are best for discriminating fetal US images without any reliance on anatomical priors or preprocessing. The generality of the learned US features makes it possible to classify the different US planes irrespective of the fetal organs. Our method achieved a mean accuracy of 94.97% and 85.74% in the classification of fetal head and heart planes, which was higher than the state-of-the-art baseline algorithms.

Keywords: Convolutional neural network, Fetal sub planes, Classification, Ultrasound images, Retrospective analysis.

1 Introduction

Ultrasound (US) imaging is the modality of choice for assessing structural and functional parameters in fetal morphology [1]. US cross-sectional analysis of each fetal organ is required to determine their development. In current clinical practice, this assessment routinely occurs by acquiring images of the fetal organs in utero and by measuring the size of different anatomical structures in the fetal anatomy [2]. As the presentation of the fetus can change during image acquisition, the reliability of the measurements obtained from the images is largely dependent upon the correct identification of defined landmarks. Recent clinical research in fetal health is attempting to determine the associations between fetal structures, which have previously been ignored in routine clinical assessments and external factors, e.g., maternal substance abuse [3], as a means of creating models to predict later health outcomes. The discovery of these associations

requires retrospective analysis on large fetal US datasets which predominantly contain 2D snapshots acquired manually by sonographers for clinical fetal monitoring. Thus it is necessary to classify the US images into different anatomical viewing planes prior to the analysis, since different anatomical structures appear on different planes. Manual categorization of these images can be tedious owing to the structural similarities shared by the planes of the same organ and time consuming to execute this task on a large number of images stored in hospital database. We suggest that an automated classification framework can be utilized to detect optimal imaging planes based on the underlying fetal organs and their specific views/planes, which will facilitate later retrospective analysis; in addition it is reproducible, and avoids inter- and intra-observer variability [4].

The automatic classification of the image planes in fetal US images is challenging due to a number of factors, such as low signal-to-noise-ratios (SNR) maternal body mass, the small size and position of the fetus. Several research studies have ventured into the classification of the fetal anatomical planes as fixed standard planes [5–10]. To support retrospective analysis Kumar et al. [5] have addressed the problem of identifying the major anatomical planes of different organs by fusing US image features with object saliency information. However, the more challenging problem is to automatically identify different planes of the same organ, which is hindered by the intra-image plane similarity. As an example, the dominant visual feature in images of the fetal cranium showing the transthalamic plane (Fig. 1 (Left) (b)) and the transcerebellar plane (Fig. 1 (Left) (c)) is the fetal skull, which obscures subtle differences between the visual appearances of different structures inside the skull. Similarly visual properties can be seen in the different planes of the heart depicted in Fig. 1 (Right).

Among the recent studies in the identification of different planes of a fetal organ, Lei et al. [6] proposed a method to classify the axial, coronal, and sagittal planes of the 2D US fetal face images, by combining scale invariant feature transform (SIFT) descriptors with an aggregated Fisher vector (FV) representation, to assist non-experts to perform accurate diagnosis and biometric measurements. Chen et al. [11] used rich feature representation of the localized fetal structures in US videos, but they required temporal information as a prior, which is not available in 2D snapshots. State-of-the-art methods for 3D US plane detection have leveraged the discriminative capabilities of machine learning technologies: Yaqub et al. [12] used Random Forests to localize the mid sagittal plane as a means of providing an accurate parasagittal plane of fetal head, while Sofka et al. [13] used an Integrated Detection Network to detect several planes in the fetal head by using the interdependence of the structural poses. However, these techniques that rely on the 3D structural information cannot be applied to classify the sub-planes of an organ in 2D US images, which are currently used in the clinical practice. Approaches for 2D US images require methods which can overcome the variability introduced during the acquisition by different machines, different operators, imaging formats, resolution, and magnification etc., and also should

not depend on any spatial (3D) or temporal priors to support the classification task.

In this paper, we propose a general framework to automatically identify the fetal sub planes in 2D US images; we demonstrate our method in the classification of the fetal sub planes of heart and head. A key element of our method is the use of convolutional neural networks (CNNs), a state-of-the-art deep learning (machine learning) technique for generating meaningful feature extractors for image data. These feature extractors can be used to derive numerical descriptors that characterize the visual properties of the image, which in turn can be used by image classification algorithms.We overcome the challenge of quantifying the subtle differences in the US images of the same organ by fine-tuning a CNN to derive an US specific feature extractor that is capable of generating image features that are tuned to distinguish US images. We achieved fetal sub plane classification using an ensemble of classifiers for each organ. To the best of our knowledge, this is the first work to report the fetal sub plane classification in 2D US images. Key contributions of our work are as follows:

1. We train an US specific feature extractor using major anatomical fetal US images, which can be utilized to represent any fetal US images.
2. We show that our method is robust and can be used for multiple organs through experiments on multiple planes of two different organs.

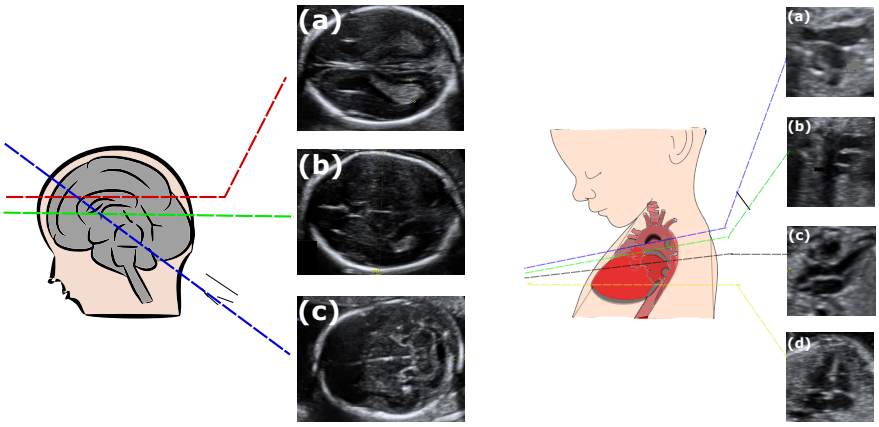

Fig. 1. Left: Axial planes of fetal head (a) Transventricular plane; (b) transthalamic plane; (c) transcerebellar plane. Right: Standardized transverse scanning planes for fetal echocardiography (a) Three vessel trachea view; (b) right ventricular outflow tract; (c) left ventricular outflow tract; (d) four chamber view.

2 Materials

We obtained a dataset of 5612 2D fetal US images (resolutions of either 640 x 480 pixels or 960 x 720 pixels). The images were acquired from 185 fetuses using a GE Voluson E8 ultrasound machine. Experienced sonographers followed the protocols defined by the AIUM [1] and International society of Ultrasound in Obstetrics and Gynecology for imaging fetal structures during 18-20 week morphology scan [14]. The images were divided into a training set (5229 images from 100 fetuses) and a test set (209 heart planes and 174 head plane images from 85 fetuses). The training images included different imaging planes of 15 major fetal structures[1].

The test images were manually classified into 4 different heart planes and 3 different head planes by a medical imaging researcher under clinical supervision to facilitate ground truth comparison. The heart planes included 77 images of four chamber (4C), 29 images of left ventricular outflow tract (LVOT), 53 images of right ventricular outflow tract (RVOT), 29 images of three vessel and tracheal view (3VT) and 21 images of heart not belonging to these planes were grouped into a single group called other heart planes. The 3 different axial head planes included 31 images of transcerebellar (TC), 28 images of transventricular (TV), 98 images of transthalamic (TT), and 17 images of other head planes. Our institutional ethics committee approved the use of these images in research.

3 US CNN Model and Feature extraction

We use CNNs to extract US image features because CNN's exploit the local correlations and sparseness in the image [15]. We trained an US specific feature detector (US CNN) by using the US dataset to fine-tune the AlexNet CNN model [16] pre-trained on the ImageNet database with 1000 natural object categories. We replaced the last fully connected layer (intended for 1000 classes) of the AlexNet architecture with a new fully connected layer for the 15 different fetal structures in our dataset. A logistic loss layer was added to facilitate the training process. The weights of the entire network of the fine-tuned model were iteratively updated using backpropagation and stochastic gradient descent. We used a learning rate of 10^{-3} for our new fully connected layer and learning rate of 10^{-4} for the rest of the network. The higher learning rate for the last layer was intended to enable faster learning of US specific weights for the 15 classes. We increased the robustness of the algorithm by augmenting the data set, with twenty four variations of the image (original crop c with additional 5 cropping from all the corners and center of c) and 3-axes flipping (x axis, y axis and both axes); thereby increasing the number of training samples. 90% of the augmented training dataset was used for training and 10% for validation.

After fine-tuning, the US images could be represented by a 4096 dimensional feature vector extracted from the last fully connected layer of the US-

[1] Major fetal structures - abdomen, arm, blood vessels, cord insert, face, femur, humerus, foot, genitals, head, heart, kidney, spine, leg and hand.

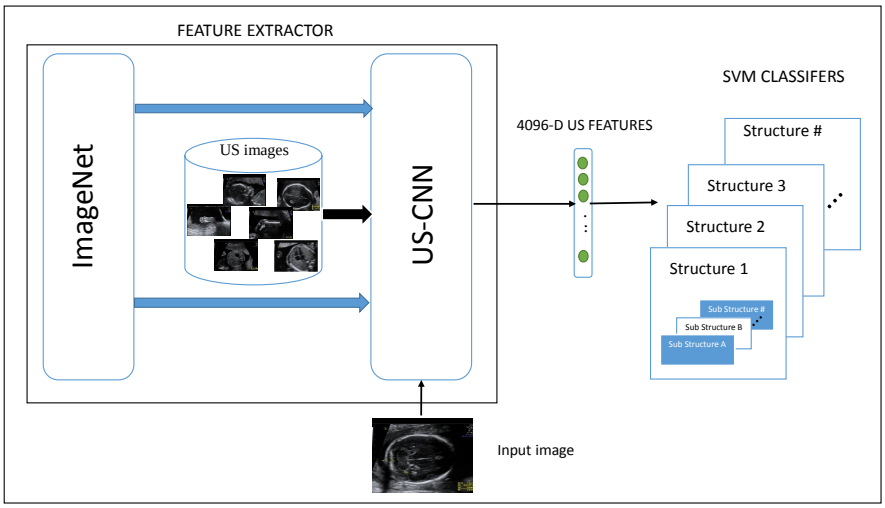

Fig. 2. The framework of Fetal sub plane classification contains a feature extractor, a classifiers for each fetal structure and sub classifiers to classify the sub-planes of the fetal structures.

CNN model. We trained a one-vs-all SVM model [17] for all of the various planes within each organ. Fig. 2 shows the overall framework of the sub plane classification.

4 Experimental Procedure

Our experiments used a 10-fold training of one-vs-all SVMs to avoid biasing the classifier to the training data. Of the 209 cardiac images, each fold used 188 images for training and 21 in testing. Of the 174 head images, each fold used 157 images for training and 17 in testing. We compared our proposed method (US CNN) with two baselines: (i) a local descriptor based approach for US plane classification (SIFT with FV encoding), as it was used in [6] to classify different planes of the same organ; and (ii) the 4096 feature vectors extracted from AlexNet without fine-tuning (i.e., trained only on 1000 classes of natural images) to study the performance of the CNN based classification. We measured the performance of these methods using the classification accuracy, precision, sensitivity, specificity and receiver operating characteristic (ROC) curves.

5 Results and Discussion

Table 1 compares our method with the baseline methods for fetal sub cranial plane classification; Table 2 compares our method with the baseline methods

Table 1. Classification results of fetal sub cranial planes

Plane	Method	Accuracy	Sensitivity	Specificity	Precision
TV	US CNN	**97.70**	**85.71**	**100**	**100**
	Baseline AlexNet	93.68	82.14	95.89	79.31
	SIFT+FV	83.33	46.42	90.41	48.14
TT	US CNN	**93.10**	**95.92**	**89.47**	**92.16**
	Baseline AlexNet	90.23	92.86	86.84	90.10
	SIFT+FV	78.73	80.61	76.31	81.44
TC	US CNN	**97.13**	**93.55**	**97.90**	**90.63**
	Baseline AlexNet	94.25	83.87	96.50	83.87
	SIFT+FV	81.03	41.93	89.51	46.42
Others	US CNN	91.95	35.29	98.09	66.67
	Baseline AlexNet	**93.68**	**52.94**	**98.09**	**75.00**
	SIFT+FV	91.37	35.29	97.45	60.00
Mean of all planes	US CNN	**94.97**	77.61	**96.36**	**87.36**
	Baseline AlexNet	92.96	**77.95**	94.33	82.07
	SIFT+FV	83.61	51.06	88.42	59.00

Table 2. Classification results of fetal sub cardiac planes

Plane	Method	Accuracy	Sensitivity	Specificity	Precision
4 Chamber	US CNN	**88.52**	**77.92**	**94.70**	**89.55**
	Baseline AlexNet	82.32	76.62	85.61	75.64
	SIFT+FV	46.41	33.77	53.77	29.89
LVOT	US CNN	80.86	17.24	91.11	23.81
	Baseline AlexNet	**84.21**	**17.24**	**95.00**	**35.71**
	SIFT+FV	68.42	3.45	78.89	2.56
RVOT	US CNN	**80.86**	**58.49**	**88.46**	**63.27**
	Baseline AlexNet	77.51	45.28	88.46	57.14
	SIFT+FV	75.11	32.08	89.74	51.52
3V	US CNN	**87.08**	**41.38**	**94.44**	**54.55**
	Baseline AlexNet	85.17	20.69	95.56	42.86
	SIFT+FV	84.68	10.34	96.66	33.33
Others	US CNN	**91.39**	**33.33**	**97.87**	**63.64**
	Baseline AlexNet	89.00	28.57	95.74	42.86
	SIFT+FV	88.99	14.28	97.34	37.50
Mean of all planes	US CNN	**85.74**	**45.67**	**93.31**	**58.96**
	Baseline AlexNet	83.64	37.68	92.074	50.84
	SIFT+FV	72.72	18.78	83.28	30.96

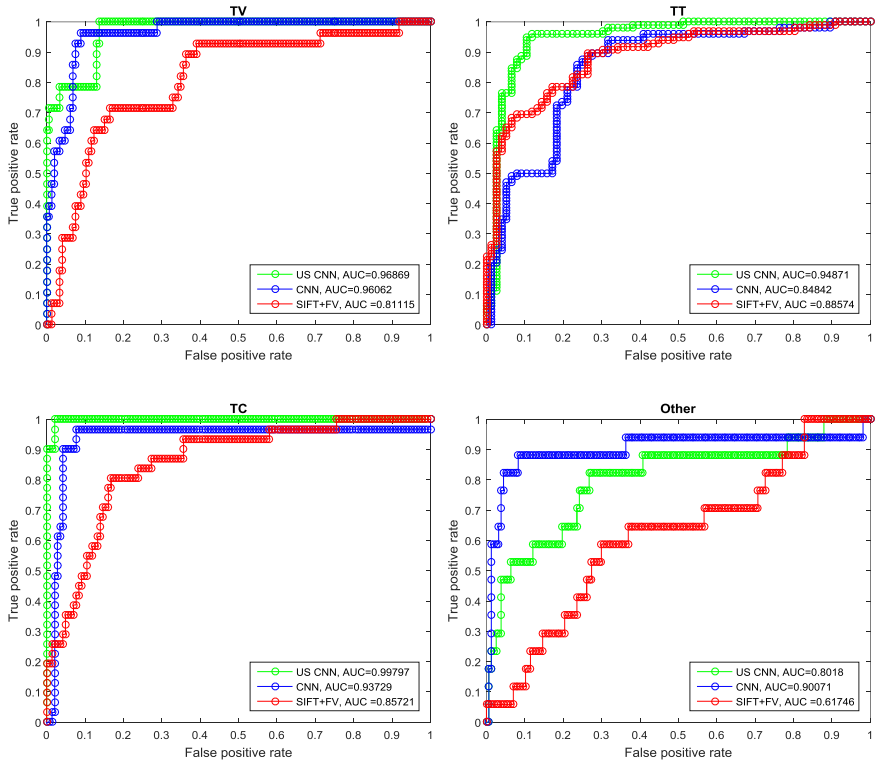

Fig. 3. ROC Curves for the detection of fetal sub cranial planes

for sub cardiac plane classification. Our method significantly outperformed the SIFT+FV and generally had a higher accuracy than the baseline AlexNet. The high accuracy of both our method and that of the baseline AlexNet confirms that CNN features are robust, even when the baseline AlexNet is trained on natural images. We attribute the high accuracy due to the extraction of the optimal features using a fine-tuned CNN for US images.

Fig. 3 shows that our method obtained the highest accuracy in all the sub-planes of the fetal head. The ROC has an area under the curve (AUC) close to 1 for all the planes of the head suggesting our method was able to discriminate the structural differences between these planes. The Other group contained a wide variety of images that did not belong to TV, TT, and TC. The variations among this group meant that they were better represented by the generic features extracted by AlexNet. We suggest that our method would have a higher accuracy if the Other images were further divided into unique subgroups prior to SVM training.

Our method was generally better than the baseline methods in sub-cardiac plane classification, especially for the four-chamber view, which is clinically im-

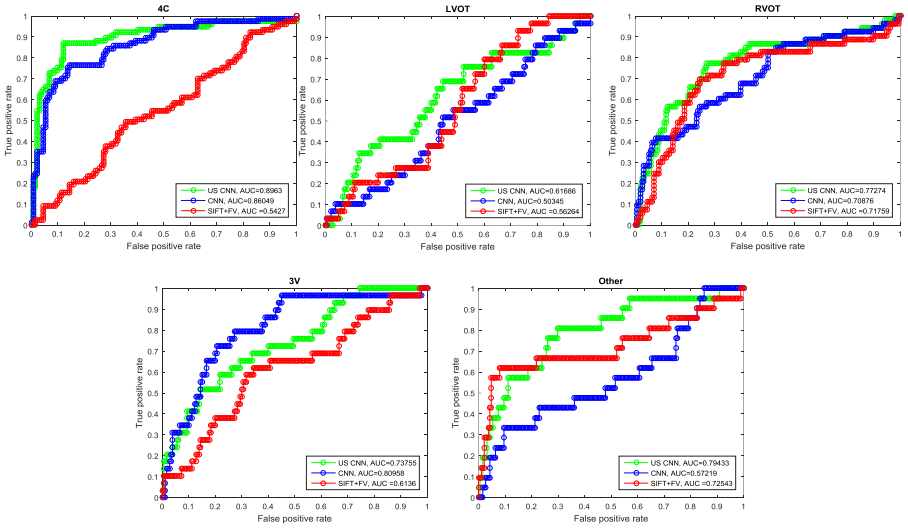

Fig. 4. ROC Curves for the detection of fetal sub cardiac planes

portant in the assessment of congenital heart diseases [18]. Our method achieved lower results for LVOT because there are many variations in these images, including images where the blood vessels appear to merge. This level of variability within a single plane makes LVOT harder to distinguish compared to planes with more consistent appearance. We can improve our methodology by expanding our training data to include more such variations, thereby improving the fine-tuning of our CNN and the training of our SVMs.

The accuracy for the heart sub-plane classification is less than that of head classification primarily because the heart sub-planes are naturally more difficult to classify. This is because the fetal heart is a dynamic organ and its appearance can dramatically change due to the contraction or expansion of the heart in addition to the motion of the fetus and angle of the ultrasound probe. In contrast, head planes generally have a static appearance and only need to recognize variations introduced by the motion of the fetus and the probe angle during acquisition. Such information can be captured from the structural poses in 3D US or the temporal deformations in US video but in 2D US, which are among the majority acquired and stored, only the specific characteristics of individual images are available.

Our results show that the well-established process of fine-tuning a pre-trained CNN to a particular task can be readily adapted to the problem of detecting different structures within a single organ and gives consistently better results than other methods. We envision structure specific improvements which can differen-

tiate dilated and narrowed blood vessels to improve the classification accuracy of outflow tracts in the heart planes. In future with organ specific optimizations, our proposed method can be extended to categorize any anatomical US images.

6 Conclusion

In this paper, we presented a method to automatically classify different structures within the same anatomical fetal organ in 2D US images. Our classification framework uses a fine-tuned CNN as feature extractor to discriminate structurally similar objects. Furthermore, our method can serve as a prior for any automated fetal segmentation, which requires accurate identification of the fetal planes. In future, we will extend the prediction to a larger dataset and investigate the hierarchical classification frame work of inter and intra fetal planes.

Acknowledgements. The authors gratefully acknowledge the support of NVIDIA Corporation through the donation of the Titan X GPU used for this research.

References

1. American Institute of Ultrasound in Medicine: AIUM practice guideline for the performance of obstetric ultrasound examinations. Journal of Ultrasound in Medicine 29(1), 157 (2010)
2. Gluckman, P.D., Hanson, M.A.: The developmental origins of health and disease. Springer (2006)
3. Schulson, M., Liu, A., Björkman, T., Quinton, A., Mann, K.P., Benzie, R., Peek, M., Nanan, R.: Mid-gestational enlargement of fetal thalami in women exposed to methadone during pregnancy. Frontiers in Surgery 1(8) (2014)
4. Sarris, I., Ioannou, C., Chamberlain, P., Ohuma, E., Roseman, F., Hoch, L., Altman, D., Papageorghiou, A.: Intra-and interobserver variability in fetal ultrasound measurements. Ultrasound in Obstetrics & Gynecology 39(3), 266–273 (2012)
5. Kumar, A., Sridar, P., Quinton, A., Kumar, R.K., Feng, D., Nanan, R., Kim, J.: Plane Identification in Fetal Ultrasound Images using Saliency Maps and Convolutional Neural Networks. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (2016)
6. Lei, B., Zhuo, L., Chen, S., Li, S., Ni, D., Wang, T.: Automatic recognition of fetal standard plane in ultrasound image. 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI) pp. 85–88 (2014)
7. Kwitt, R., Vasconcelos, N., Razzaque, S., Aylward, S.: Localizing target structures in ultrasound video–a phantom study. Medical Image Analysis 17(7), 712–722 (2013)
8. Yaqub, M., Kelly, B., Papageorghiou, A.T., Noble, J.A.: Guided Random Forests for Identification of Key Fetal Anatomy and Image Categorization in Ultrasound Scans. In: MICCAI 2015, pp. 687–694. Springer (2015)
9. Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.A.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE Journal of Biomedical and Health Informatics 19(5), 1627–1636 (2015)

10. Carneiro, G., Georgescu, B., Good, S., Comaniciu, D.: Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Transactions on Medical Imaging 27(9), 1342–1355 (2008)
11. Chen, H., Dou, Q., Ni, D., Cheng, J.Z., Qin, J., Li, S., Heng, P.A.: Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks. In: MICCAI 2015, pp. 507–514. Springer (2015)
12. Yaqub, M., Rueda, S., Kopuri, A., Melo, P., Papageorghiou, A., Sullivan, P., McCormick, K., Noble, J.: Plane localization in 3d fetal neurosonography for longitudinal analysis of the developing brain. IEEE Journal of Biomedical and Health Informatics PP(99), 1–1 (2015)
13. Sofka, M., Zhang, J., Good, S., Zhou, S.K., Comaniciu, D.: Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and integrated detection network. IEEE Transactions on Medical Imaging 33(5), 1054–1070 (2014)
14. International Society and Others of Ultrasound in Obstetrics & Gynecology Education Committee: Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. Ultrasound in Obstetrics & Gynecology 29(1), 109 (2007)
15. Li, J., Zhang, T., Luo, W., Yang, J., Yuan, X.T., Zhang, J.: Sparseness analysis in the pretraining of deep neural networks. IEEE Transactions on Neural Networks and Learning Systems PP(99), 1–14 (2016)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems pp. 1–9 (2012)
17. Rifkin, R., Klautau, A.: In defense of one-vs-all classification. The Journal of Machine Learning Research 5, 101–141 (2004)
18. Copel, J.A., Pilu, G., Green, J., Hobbins, J.C., Kleinman, C.S.: Fetal echocardiographic screening for congenital heart disease: the importance of the four-chamber view. American Journal of Obstetrics and Gynecology 157(3), 648–655 (1987)