

A Deep Convolutional Neural Network based Hybrid Framework for Fetal Head Standard Plane Identification

Jingyu Ye¹, Ruizhi Liu², Bin Zou³, Hongyang Zhang², nianji zhan⁴, Cong Han², Ying Yang¹, Hongguo Zhang², Jian Guo¹, Fang Chen¹, Shida Zhu¹, and Shucheng Hua⁵

¹BGI-Shenzhen

²Jilin University First Hospital

³CNGB

⁴Affiliation not available

⁵Jilin University First Hospital Department of Respiratory Medicine

May 6, 2020

Abstract

As considered to be less risky, less expensive, and more convenient than radiological examinations, ultrasound has been routinely employed in prenatal exams for the past decades. However, the quality of acquired ultrasound samples, i.e., ultrasound images or videos, and the further diagnosis is crucially depended on the sonographer. At the meantime, there are an extremely limited number of experienced sonographer available for the fetal ultrasound screening. Therefore, to reduce the workload of sonographers, and to promote the quality of fetal ultrasound screening, a deep convolutional neural network based framework is proposed for automatically differentiating five types of fetal head ultrasound standard planes, i.e., Transventricular plane (TV), Transthalamic plane (TT), Transcerebellar plane (TC), Coronal view of eyes (Eyes), Coronal view of nose (Nose), and other non-standard fetal head ultrasound images (Background). A dataset consists of 19928 fetal ultrasound images is applied for the model training and performance evaluation. By combining object detection network, object classification network, and model stacking technique, the proposed framework achieves the state-of-the-art performance with the average accuracy of 89.61% and the average F-1 score of 89.61%.

A Deep Convolutional Neural Network based Hybrid Framework for Fetal Head Standard Plane Identification

Jingyu Ye^{1*}, Ruizhi Liu^{2*}, Bing Zou^{1*}, Hongyang Zhang², Nianji Zhan¹, Cong Han², Ying Yang^{1,3}, Hongguo Zhang², Jian Guo¹, Fang Chen^{1,3}, Shida Zhu^{1#}, Shucheng Hua^{2#}

¹ BGI-Shenzhen, Shenzhen 518083, China; ² Reproductive Medicine & Prenatal Diagnosis Center, The First Hospital, Jilin University, Changchun, China.; ³ Shenzhen Engineering Laboratory for Birth Defects Screening, BGI-Shenzhen, Shenzhen 518083, China; *These authors contributed equally; #Correspondence author: Shucheng Hua, The First Hospital, Jilin University, No.71 Xinmin Street, Changchun 130021, China. Tel. +86043188782707. Email: shuchenghua@126.com; Shida Zhu, BGI-Shenzhen, Building 11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China. Tel. +8615814003823. Email: zhushida@bgi.com

Abstract :

Objective To develop more effective deep convolutional neural network based technology for automatic identify five types of fetal head standard view planes, i.e., Transventricular plane (TV), Transthalamic plane (TT), Transcerebellar plane (TC), Coronal view of eyes (Eyes), Coronal view of nose (Nose), and other non-standard fetal ultrasound images (Background) in ultrasound images.

Design Randomly select samples for the research including the abnormal fetuses.

Setting Historical records of prenatal examinations from The First Hospital of Jilin University.

Samples 19928 2-D fetal ultrasound images of fetuses with the gestational age ranging from 18 weeks to 31 weeks.

Methods A novel convolutional neural network based hybrid framework is designed for automatic identification of standard fetal head view planes in ultrasound images which is an important topic of fetal screening and diagnosis. In the proposed framework, YOLO-V3 is applied for locating possible fetal head region in the candidate image. Then, a group of object classification networks which includes ResNet50, ResNeXt50, InceptionResNet-V2, and SonoNet64 are employed to make predictions on the located fetal head region. The predictions of the classification models are stacked so as to generate the final prediction for the suspected fetal head region, and the original image as well.

Main Outcome Measures The average precision, average recall, average F-1 score, and the average accuracy on identifying six categories of fetal ultrasound images for each method.

Results The proposed deep convolution neural network based framework achieves the average precision of 89.67%, the average recall of 89.61%, the average F-1 score of 89.61%, and the average accuracy of 89.61% which demonstrates the state-of-the-art performance. The average AUC is 0.9893.

Conclusions The experimental results indicate that the proposed framework is effective on identifying fetal head standard view planes in ultrasound images. Since the experiments are designed to reproduce the scenario happened in the real life, the proposed method could be potentially applied to the automatic fetal screening and diagnosis.

Keywords : fetal ultrasound screening, deep convolutional neural network, fetal head standard plane, model stacking.

I. Introduction

It is extremely crucial for the parents to know if their babies are healthy or not during the pregnancy. The earlier the fetal abnormality could be found, the more chance the abnormality could be cured. Therefore, increasing attention and tremendous efforts have been put on improving the harmlessness, effectiveness, and robustness of fetal abnormality screening by researchers.

Although radiological examinations are known to be more precise on revealing abnormalities of human body, the side-effects are also significant. To avoid such potential, threaten to the fetus and the mothers, a less risky technology is required for the screening. Thus, ultrasound which has been proved to be less harmfulness, less expensive, and more convenient than other radiological techniques is widely utilized for the fetal abnormality screening. However, ultrasound examination is less automatic than other radiological exams as the sonographer needs manually hold the probe to conduct the examination. Therefore, more experienced and skillful sonographers are required for the ultrasound screening otherwise the results could not be reliable. Unfortunately, the shortage of experienced sonographers has been a severe problem ever since the born of ultrasound, and a qualified sonographer may need years for training which is extremely time consuming. To solve these problems, computer aided technologies are urgently demanded.

In the screening, the priority for a sonographer is to find a series of clinical standard planes which are a set of anatomical views of the fetus. Then, the diagnosis is to be made based on both subjective evaluation which is the observation of the sonographer, and objective evaluation which includes several physical measurements. Obviously, the final diagnosis is highly depending on the quality of standard anatomical views acquired by the sonographer. In other words, as far as the standard fetal anatomical planes can be precisely obtained, the diagnosis is to be more accurate and more reliable.

For the retrospective studies of the fetal growth and diseases, well organized and categorically arranged retrospective data are always required. However, that could usually be not true since all of the historical

data are stored without any appropriate arrangement for most of the hospitals. Therefore, a technique which can automatically differentiate and separate the historical data according to the clinical meaning or usage is to be beneficial to the researches.

In this work, a deep neural network based framework is presented for classifying various types of standard anatomical planes of fetal head, i.e., Transventricular plane (TV), Transthalamic plane (TT), Transcerebellar plane (TC), Coronal view of eyes (Eyes), Coronal view of nose (Nose), and other non-standard fetal ultrasound images (Background). In the proposed framework, a YOLO based object detection network is applied to locate the head region in ultrasound image, and a set of powerful classification networks are utilized with model stacking technique to give a final judgement on each image based on the detected head regions. The contributions of this work are as follows: first of all, this is the first piece of work using deep learning technology to identify TV, TT, and TC to the best knowledge; secondly, this is the first piece of work which successfully applied YOLO on this topic; thirdly, the design of combining object detection network, object classification network, and model stacking technique is novel to this area; finally, the proposed framework achieves the state-of-the-art performance.

The rest of this paper is arranged as follows. In Sec. II, related works are briefly introduced. Then, the proposed framework is to be presented in Sec. III. Experimental results and discussion are reported in Sec. IV. The conclusion is finally made in Sec. V.

II. Related works

Artificial neural network(1), which was addressed in last 50s, is one type of network topologically composed of multiple artificial ‘neurons’ so as to mimic the way how human being learns and acts. In last 80s, convolutional neural network was designed for computer vision tasks such as hand writing digits recognition. By the enormous development of computing hardware and the boost of the Internet, multiple layers’ CNN(2, 3), also known as deep CNN, was pushed in front of the stage in the past decade due to its extraordinary performance achieved in computer vision contests.

To automatically distinguish the standard fetal anatomical scan planes, several works have been presented for 2-D ultrasound images or videos. In(4), Active Appearance Models are utilized to identify if the composite structure of butterfly-shaped thalami and the falx is appeared in the scan planes, and a score function is applied for evaluating the correctness of the planes with detected structure. (5)proposed to use a hybrid model which composites of both convolutional neural network (CNN) and long short term memory(6) (LSTM) model to locate fetal standard planes in ultrasound videos. In(7), a CNN is proposed for identifying fetal abdominal standard planes in ultrasound videos. A fisher vector based model is presented in (8) for the recognition of fetal facial standard planes in ultrasound images. A CNN based model which is called SonoNet(9) is presented for the real-time detection of fetal standard planes while the deepest SonoNet involves 13 convolutional layers. In (10), a 16 convolutional layers based CNN is suggested to be able to recognize three types of fetal facial standard planes. In(11), the information extracted from both cropped regions of fetal structures and the whole ultrasound image via CNN are suggested to be fused in order to identify fetal standard planes. Besides, a few works have been carried out for the localization of fetal standard planes in 3-D ultrasound volumes(12, 13).

III. Methodology

The process of fetal abnormality screening normally consists of two steps: (1) looking for standard planes, (2) evaluation and diagnosis based on standard planes. When the sonographer sweeping the ultrasound tube and looking for the standard planes of fetal bodies, it is actually object detection discussed in computer vision. In order to discriminate fetal standard planes, a deep neural network based hybrid framework is proposed. The idea behind the proposed framework is to use an object detector to locate the Region of Interests (ROI), i.e., fetal head region, then another network is applied to make further decision on the extracted ROI. In this work, YOLO-V3(14) is applied as the object detector, and ResNeXt(15) is employed as the fine-grained classifier.

Object Detection Network

Object detection, particularly with deep neural networks (DNNs) instead of hand-crafted features, has been addressing immense attention in the past decade and tremendous development has been delivered. Generally, two types of DNN based object detection frameworks are dominating, i.e., (1) one-stage networks, e.g., all variants of YOLO(16) and SSD(17); (2) two-stage networks, e.g., all variants of RCNN such as Faster RCNN(18). The major difference between one-stage and two-stage networks is the utilization of Region Proposal Network (RPN)(18). In one-stage networks, both localization and classification are conducted simultaneously within the same network, while RPN are employed specifically for object localization in parallel with the feature extraction in two-stage networks. Therefore, two-stage networks are normally more precise on object localization especially for small objects, but more time and computing resource consuming on the training stage and the inference stage as well. In the proposed framework, the object detection network is utilized to locate the fetal head related region in the candidate sources, i.e., ultrasound images, and predict the corresponding label for the candidate based on the detected image content. Later on, the located fetal head region is to be cropped from the source image and further input into the followed fine-grained classification network to get the final prediction of the source image. In this work, YOLO-v3 with Darknet-53 (14) as the backbone of feature extractor is employed for locating fetal head region in the proposed framework.

YOLO has been proved its efficiency in enormous prior works. In order to elevate the performance of vanilla YOLO, more powerful backbone and more advanced design are included in the architecture of the latest version of YOLO which is named as YOLO-V3. The whole network of YOLO-V3 can be considered as the composition of two subnetworks, Darknet-53 for the feature extraction and Feature Pyramid Network (FPN) for multi-scale object detection. In Darknet-53, the idea of ResNet (19) is comprehensively injected. Multiple residual modules (RMs), which are 23 RMs in total, are employed and divided into five groups. Each RM is composited with two convolutional layers with different filter sizes, i.e., 1×1 and 3×3 . After each RM group but not the last RM group, a convolutional layer with the stride as 2 is connected for down-sampling. Therefore, the dimensionality of the output of the down-sampling layer, i.e., height x width, is reduced to 1/4 of the feature map produced by the previous RM group. Also noted that, each convolutional layer is followed by a Batch normalization layer and a non-linear activation layer equipped with leaky ReLU(20). Batch normalization (BN) (21) layer is designed to diminish the internal covariate shift (21) which could easily lead gradient vanishing in the training of deep neural networks. Besides, BN introduces regularization to the network which enhances the generalization of the whole network. Therefore, the utilization of BN makes the training of neural networks particularly extremely deep neural networks become possible. Similarly, leaky ReLU is able to reduce the risk of gradient vanishing occurred in back propagation of deep neural networks as well thus to accelerate the training process. More importantly, non-linearity allows networks to model more complicated problems such as most of real-life problems.

In YOLO-V3, FPN is also applied in order to improve the performance of previous YOLOs. By investigating the multi-scale feature representations, which are the output of layers of the feature pyramid, the network could find the most accurate locations of objects in variate scales, and thus determine what they are. Three different scales are considered in the FPN of YOLO-V3, while three candidate boxes which are representing the possible locations of the detected object are predicted at each scale. For each box, a set of parameters, which includes the coordinates of the centroid and the offsets of the width and the height, is to be generated by the network so as to indicate a certain region in the original image. Besides, a confidence score and the class predictions are also produced for each predicted bounding box.

2. Object Classification Network and Model Stacking

Since the invention of more advanced computing hardware and the explosive growth of the data, more complicated neural networks are allowed to be built and trained. However, it is noticed that a deeper network not always gains a better performance but a higher error rate and a degradation of the accuracy. To resolve this problem, the idea of residual learning is brought by ResNet(19) which is considered as one of the most significant progress on deep learning. By utilizing identity mapping (22), also known as residual skipping

connection, information can easily pass through even between the deeper layers. At the meantime, the residual connection could effectively avoid gradient vanishing even weights are extremely small. Therefore, deeper neural networks with RM are much easier to get converged and achieve a better result, so as YOLO-V3. Based on this RM technique, ResNet, InceptionResnet(23), ResNeXt(15) are proposed.

ResNet as one of the most popular CNN based network is designed to overcome the difficulty of training deep neural networks. Before ResNet is proposed, a deeper neural network is hard to train and the degradation of the performance is always happened. The problem comes from the gradient vanishing occurred in the training. By introducing the residual short cut, which is the essential idea of RM, the gradient can easily flow in the network during the back-propagation. Consequently, the efficiency of training is dramatically improved and the performance is boosted. Inception is another milestone. Not only the depth of the network is increased in Inception, but also the width is extended as well. It tries to extract various types of features from the same source (image or feature maps) so as to enhance the ability of the network. With injecting the idea of ResNet, InceptionResNet-V2 is developed and more successful results are achieved. In ResNeXt, multiple RMs are successively stacked which is based on the idea of VGG(24) in order to increase the depth of the whole network. Besides, not only could the depth dramatically influence the performance of a DNN, but the width and the cardinality are both essential to the DNN as well. Considering all these factors, the idea behind Inception is also adopted by ResNeXt. For each RM, the single pathway is replaced by multi-branch topology which is also known as split-transform-merge structure. The hyperparameters of each branch within a multi-branch RM is set to be identical. This modified RM is more powerful than the vanilla single branch RM. Comparing with complicated dense structures, the modified multi-branch RM is achieved very close performance while it has much lower computational complexity.

After the inference of YOLO-V3, all possible fetal head regions within candidate images are assumed to be located. Instead of relying on the classification of YOLO-V3, all detected regions are sent to a set of powerful networks specifically designed for the object classification. The final prediction of each suspected image regions is to be made based on the stacked outputs of the classification networks so as the corresponding original image.

IV. Experimental Results

To further evaluate the proposed framework, 19928 2-D fetal ultrasound images collected by the center of prenatal diagnosis of The First Hospital of Jilin University are utilized to form the experimental dataset. Images are from the historical screening records of 2913 volunteers (cases) who has taken the standard ultrasound prenatal examination in The First Hospital of Jilin University from August 2017 to March 2018. The gestational age of fetuses is ranging from 18 weeks to 31 weeks. Examinations are conducted by a group of experienced sonographers and GE Volution E8 ultrasound scanners are employed as the screening equipment. In this project, volunteers are anonymous and personal information has been removed from each image. Besides, the data collection is fully under the supervision of the ethics committee of the hospital.

All images are labelled into six categories, i.e. five types of fetal head standard view planes and the background, by a set of experienced sonographers. The standard planes are Transventricular plane (TV), Transthalamic plane (TT), Transcerebellar plane (TB), Coronal view of nose (Nose), Coronal view of eyes (Eyes), respectively. The background includes other types of fetal ultrasound images, i.e. fetal head images but not the standard views required by the screening and abdominal views. The distribution of the images is illustrated in Figure 2. The images are divided to three portions with the ratio of 8:1:1 for training, validation, and testing, respectively. Since each case would have multiple images involved, the case included in the training or validation set is excluded from the testing set in order to avoid data leakage. In the training stage, both training and validation sets are utilized, while testing set is only applied for the performance evaluation.

To train the YOLO, each image with the label, and the four coordinates of its fetal head region, i.e., top-left corner and bottom-right corner, are fed into the network. Here, only the image with fetal head region is utilized for the training no matter it is standard head plane or background since the number of images without fetal head region is very limited and the fetal body lies in such background image is not the interest

of this work. The YOLO architecture used in this work is designed for images with the size of 320×320 (width \times height), therefore each image is resized to 320×320 before input into the YOLO while the batch size for training is set to 16 images. The performance of YOLO is evaluated on the testing set, and the evaluation is illustrated as Table 1. The last row representing the average values of precision, recall, and F1-score of the six categories, respectively.

As illustrated in Figure 1, the local regions where fetal heads are detected are cropped from the original images, such as the first two images in which the head regions are found by YOLO. For background images which no fetal head is shown, such as the third sample shown in Figure 1, a pre-set rectangle window is applied so as to extract the main image content from this type of images. The reason behind is to preclude the fetal head images which could possibly missed by YOLO. New samples are resized to 224×224 for ResNet50, ResNeXt50, and SonoNet64, and 299×299 for InceptionResNet-V2. The settings for training of the aforementioned series of networks are similar. Transfer learning is conducted based on the models pre-trained on ImageNet. Adam is employed as the optimizer while the initial learning rate is set to 0.001. The learning rate is reduced by the factor of 0.1 if the validation loss is not improved during the past 10 epochs. Cross-entropy is applied as the loss function, and the model with the smallest validation loss is selected for each type of classification networks when the training process is accomplished. In this work, the categorical predicted probabilities of each sample generated by each model are utilized as for followed model stacking instead of the predicted labels. Therefore, 24 probabilities (six classes \times 4 models) can be obtained for each sample. Then, the model stacking is conducted based on the following equation,

$$\tilde{Y}_i = \operatorname{argmax}_c (P_{i,c}),$$

$$P_{i,c} = \left\{ \frac{1}{m} \sum_m p_{i,m,1}, \dots, \frac{1}{m} \sum_m p_{i,m,c} \right\},$$

where $p_{i,m,c}$ is the probability of sample I belongs to class c predicted by model m , $P_{i,c}$ is the set of categorically averaged probabilities generated by m different models, and \tilde{Y}_i is the final predicted label of sample i . The performance of each individual network is evaluated on the testing set, and the stacking model as well. The precision, recall, F1-score, and accuracy averaged over six classes are employed as the metrics of evaluation. As reported in Table 2, the proposed stacking model outperforms each individual network.

Discussion

In this work, a hybrid framework is proposed for discriminating five types of standard fetal head view planes in ultrasound images and the background images. The contributions of this work can be summarized as the following four points. 1) It is the first work that deep learning technology is utilized to identify TV, TT, and TC to the best knowledge. 2) It is the first piece of work which YOLO-V3 is successfully applied on this topic. 3) The design of the proposed framework which contains both object detection network and object classification network while model stacking technique is applied to further enhance the performance is novel to this area. 4) The overall performance of the proposed framework is the state-of-the-art.

Considering the combination of object detection network and object classification network, one question would be intuitively raised, i.e., why both networks are needed to determine whether a given fetal ultrasound image is one of the five types of standard view planes or none of them? Firstly, it is known that object detection network is one type of multi-task network which can achieve both object localization and object classification. However, the optimization of both tasks are simultaneously conducted which means it could be more difficult than single tasking optimization, and the trade off between the two tasks could make the network less effective than single tasking network on the single task. Secondly, since the ultrasound image is always noisy, it is necessary to extract the most important information which is the most significant image content from the ultrasound image so as to make the classification more efficient. Therefore, both object detection network and classification network are deployed to from the proposed hybrid framework. To further

verify the aforementioned, a set of experiments are designed. Both SonoNet64 and ResNet50 are trained with the original ultrasound images instead of the fetal head region detected by YOLO. As reported in Table 3, the performance of both networks which are trained with original images is much lower than the networks trained with the image content extracted by YOLO. It is also observed that YOLO itself is underperformed by both SonoNet and ResNet trained with the detected fetal region. As a conclusion, the combination of both object detection network and classification network is more effective than any single type of them. Besides, as the classification network is employed, YOLO can be more focused on locating possible fetal region in the candidate image. Therefore, the classification loss of YOLO is multiplied a factor of 0.5 so as to emphasize the optimization of the object localization during the training. By investigating the testing results, the region where fetal head is captured in each standard view plane image can be located by YOLO successfully.

As illustrated in Table 2, the performance of SonoNet64, the benchmark technique applied to identify fetal ultrasound standard planes which is inspired by VGG16, is competitive with the other three benchmark neural networks which are extremely popular in Computer Vision. Since all these CNN based techniques are powerful enough on feature extracting, and it is noticed that the complexity of the network is not always proportional to the performance, such as InceptionResNet-V2 is underperformed by other three. Therefore, to further improve the performance of the state-of-the-art, model stacking is considered to be one of the solutions. By reviewing the results shown in Table 3, it is observed that the proposed stacking model outperforms the four benchmarks. The Receiver Operating Characteristic (ROC) curves shown in Figure 3 also indicate the effectiveness of proposed stacking while each AUC is close to 1 and the average AUC is 0.9893.

V. Conclusion

In this work, a hybrid framework is designed for fetal head standard plane detection. The proposed framework composites of two parts, i.e., object detection network and object classification network. With the power of two networks, the designed framework is able to identify five types of standard planes with satisfactory performance. By introducing proposed model stacking, the performance of the proposed framework can be further improved. Comparing with the state-of-the-art of fetal ultrasound standard plane classification, i.e., SonoNet64, the average accuracy has been boosted to 0.8961. The average AUC is 0.9893 which is also indicated the effectiveness of the proposed hybrid framework. Since the experiments are designed to reproduce the scenario happened in the real life, the proposed method could be potentially applied to the automatic fetal screening and diagnosis.

Acknowledgement

We appreciate all doctors who participated in the data acquisition and data labelling. We also appreciate the support from China National Gene Bank.

Disclosure of Interests

The authors declare that they have no competing interests.

Contribution to Authorship

S. Z. and S. H. initiated this project. R. L., H. Z., C. H., H. Z. and S.H. acquired and labeled the data. J.Y., designed the experiments. J. Y., B. Z., and N. Z. performed the experiments. J. Y., B. Z., Y. Y., and R. L. analyzed the data and the experimental results. J. Y., R. L., B. Z., J. G., F. C., S. Z., and S. H. wrote the paper. All authors reviewed the manuscript.

Details of Ethics Approval

The ethics approval of the whole study was granted on August 2nd 2018 by the Medical Ethics Committee of First Hospital of Jilin University (No. 2018-398).

Funding

This study was supported by Science, Technology and Innovation Commission of Shenzhen Municipality under grant No. JCYJ20180703093402288, Department of Science and Technology of Guangdong Province under grants No. 2019B020227001, Finance Department Health Special Project of Jilin Province under grants No. JLSCZD2019-022.

Reference

1. Kleene SC. Representation of events in nerve nets and finite automata. RAND PROJECT AIR FORCE SANTA MONICA CA; 1951.
2. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural computation*. 1989;1(4):541-51.
3. LeCun Y. Generalization and network design strategies. *Connectionism in perspective*. 19: Citeseer; 1989.
4. Liu X, Annangi P, Gupta MD, Yu B, Padfield DR, Banerjee J, et al. Learning-based scan plane identification from fetal head ultrasound images. *Proceedings of SPIE*. 2012;8320.
5. Chen H, Dou Q, Ni D, Cheng J-Z, Qin J, Li S, et al. Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science 2015*. p. 507-14.
6. Sak H, Senior AW, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014.
7. 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 Journal of Biomedical and Health Informatics*. 2015;19(5):1627-36.
8. Maurits NM, Lei B, Tan E-L, Chen S, Zhuo L, Li S, et al. Automatic Recognition of Fetal Facial Standard Plane in Ultrasound Image via Fisher Vector. *Plos One*. 2015;10(5):e0121838.
9. 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.
10. Yu Z, Tan E-L, 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 Journal of Biomedical and Health Informatics*. 2018;22(3):874-85.
11. Sridar P, Kumar A, Quinton A, Nanan R, Kim J, Krishnakumar RJUim, et al. Decision Fusion-Based Fetal Ultrasound Image Plane Classification Using Convolutional Neural Networks. 2019.
12. Sofka M, Zhang J, Good S, Zhou SK, Comaniciu D. Automatic Detection and Measurement of Structures in Fetal Head Ultrasound Volumes Using Sequential Estimation and Integrated Detection Network (IDN). *IEEE Transactions on Medical Imaging*. 2014;33(5):1054-70.
13. 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 in medicine & biology*. 2017;43(1):286-300.
14. Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv preprint arXiv:180402767*. 2018.
15. Xie S, Girshick R, Dollár P, Tu Z, He K, editors. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017.
16. Redmon J, Divvala SK, Girshick RB, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *computer vision and pattern recognition*. 2016:779-88.
17. Liu W, Anguelov D, Erhan D, Szegedy C, Reed SE, Fu C, et al. SSD: Single Shot MultiBox Detector. *European conference on computer vision*. 2016:21-37.

18. Ren S, He K, Girshick RB, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39(6):1137-49.
19. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *computer vision and pattern recognition*. 2016:770-8.
20. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:150500853*. 2015.
21. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *international conference on machine learning*. 2015:448-56.
22. He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. *european conference on computer vision*. 2016:630-45.
23. Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. 2016.
24. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *international conference on learning representations*. 2015.

Hosted file

Figure 1.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>

Hosted file

Figure 2.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>

Hosted file

Figure 3.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>

Hosted file

Table 1.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>

Hosted file

Table 2.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>

Hosted file

Table 3.docx available at <https://authorea.com/users/317232/articles/447374-a-deep-convolutional-neural-network-based-hybrid-framework-for-fetal-head-standard-plane-identification>