ORIGINAL ARTICLE

Computer-aided diagnosis for fetal brain ultrasound images using deep convolutional neural networks

Baihong Xie[1] · Ting Lei[2] · Nan Wang[3] · Hongmin Cai[1] · Jianbo Xian[1,3] · Miao He[2] · Lihe Zhang[2] · Hongning Xie[2]

Abstract

Purpose Fetal brain abnormalities are some of the most common congenital malformations that may associated with syndromic and chromosomal malformations, and could lead to neurodevelopmental delay and mental retardation. Early prenatal detection of brain abnormalities is essential for informing clinical management pathways and consulting for parents. The purpose of this research is to develop computer-aided diagnosis algorithms for five common fetal brain abnormalities, which may provide assistance to doctors for brain abnormalities detection in antenatal neurosonographic assessment.

Methods We applied a classifier to classify images of fetal brain standard planes (transventricular and transcerebellar) as normal or abnormal. The classifier was trained by image-level labeled images. In the first step, craniocerebral regions were segmented from the ultrasound images. Then, these segmentations were classified into four categories. Last, the lesions in the abnormal images were localized by class activation mapping.

Results We evaluated our algorithms on real-world clinical datasets of fetal brain ultrasound images. We observed that the proposed method achieved a Dice score of 0.942 on craniocerebral region segmentation, an average F1-score of 0.96 on classification and an average mean IOU of 0.497 on lesion localization.

Conclusion We present computer-aided diagnosis algorithms for fetal brain ultrasound images based on deep convolutional neural networks. Our algorithms could be potentially applied in diagnosis assistance and are expected to help junior doctors in making clinical decision and reducing false negatives of fetal brain abnormalities.

Keywords Fetal brain abnormalities · Deep convolutional neural network · Craniocerebral segmentation · Prenatal ultrasound images · Computer-aided diagnosis

Introduction

Fetal brain abnormalities are some of the most common congenital malformations. Long-term follow-up studies suggest that the incidence of intracranial abnormalities may be as high as one in 100 births [1]. As a non-invasive, non-radiative, convenient and dynamic observation approach,

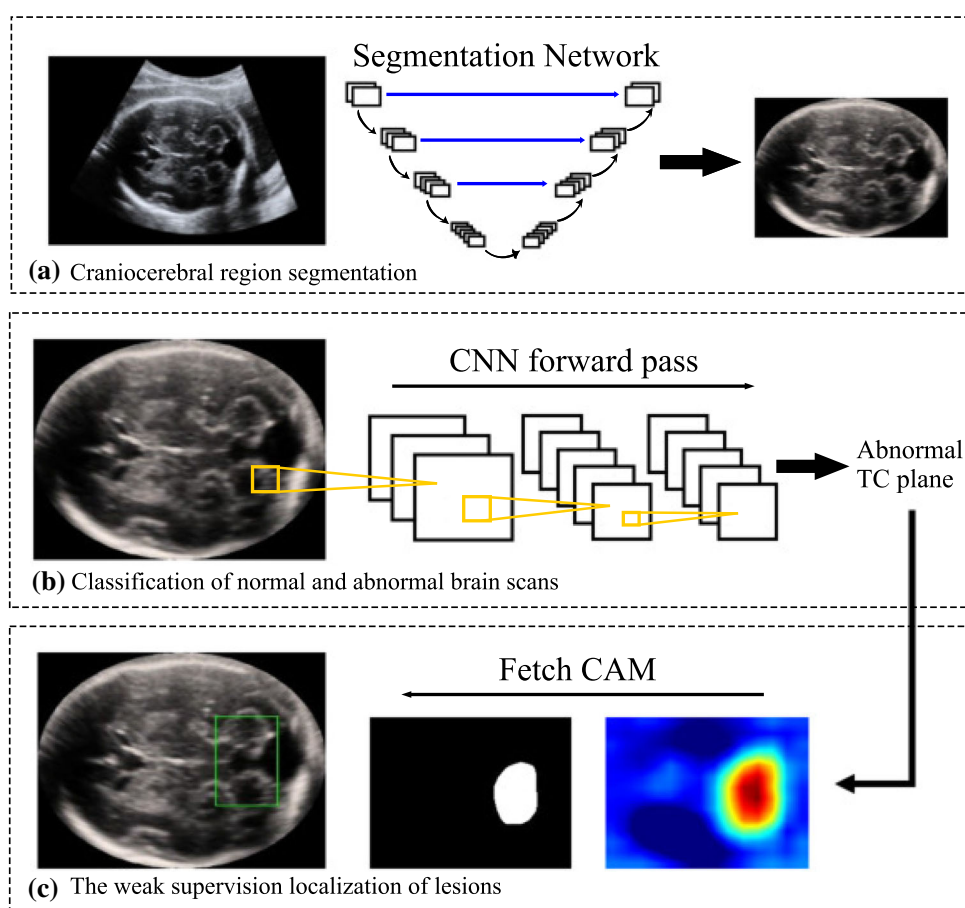Baihong Xie and Ting Lei have contributed equally to this work.

✉ Hongning Xie
  hongning_x@126.com

[1] South China University of Technology, Guangzhou, China

[2] Department of Ultrasonic Medicine, Fetal Medical Center, First Affiliated Hospital of Sun Yat-sen University, Zhongshan Er Road 58, Guangzhou 510080, Guangdong, China

[3] Guangzhou Aiyunji Information Technology Co., Ltd., Guangzhou, China

transabdominal sonography is the first choice for diagnosing fetal brain diseases. To assess the anatomic integrity of the brain, two standard axial neurosonographic planes (SANPs), transventricular (TV) and transcerebellar (TC), that allow for visualization of cerebral structures are considered [2].

Ultrasound has a long history of detecting fetal brain abnormalities. The detection rate is impaired by doctors' lack of experience in complex brain anatomy and pathology mostly [3,4], as well as outdated equipment, inappropriate fetal head position, early or late gestational age, maternal obesity. Doctors need long-time practice to be experts. However, even in expert hands some types of anomalies may be difficult to diagnose in ultrasound [1]. The development of methods that assist in prenatal sonographic diagnosis has been lacking. We think that automatic methods for diagnosing fetal congenital diseases could ease the prevalent situation of expert shortage in underdeveloped areas and could have an important clinical value.

Fig. 1 Overview of the proposed algorithms: **a** the craniocerebral regions are segmented out from 2D fetal ultrasound images by fully convolutional network (as long as the images contain a circular skull shape); **b** the segmented images are processed by the convolutional neural networks to diagnose the plane as normal or abnormal (in this case, the abnormal transcerebellar (TC) plane is diagnosed); **c** if an abnormal plane is predicted, the lesions are localized by fetching class activation mapping (CAM) from the networks

(a) Craniocerebral region segmentation

(b) Classification of normal and abnormal brain scans

(c) The weak supervision localization of lesions

In recent years, deep learning has emerged as a promising technique. Several studies on recognition [5,6], detection [3,7,8] and localization [3,8] of fetal sonographic standard planes have been reported. However, these studies did not consider abnormal cases as we do in our research. In this paper, we chosen fetal brain abnormalities as the focus of automatic diagnosis due to their high rate of occurrence.

Contributions

We introduce deep learning algorithms for computer-aided diagnosis of fetal brain abnormalities. The algorithms robustly segment the craniocerebral regions, accurately classify fetal brain images of SANPs as normal or abnormal and locate the lesions to provide visualized evidence for the diagnosis. An overview of the proposed methods is shown in Fig. 1. To the best of our knowledge, our study is the first attempt to develop algorithms that determine whether fetal brain images of SANPs are normal or abnormal.

Related studies

Standard planes studies Several reported studies are based on extracting Haar-like features from the data and training clas-

sifiers such as AdaBoost or random forests [9,10]. Recently, the trend of using convolutional neural networks (CNNs) to analyze ultrasound images has gradually emerged due to advances in the field of computer vision. The most related studies are from Yaqub [6] and Baumgartner [3]. Yaqub [6] designed networks with four convolutional layers and one fully connected (FC) layer to detect cavum septi pellucidi (CSP) structure on the TV planes. Baumgartner [3] trained networks based on VGG-net [11] to detect 13 kinds of standard planes including the TV and TC planes. Both studies report that the images of SANPs could be recognized by CNN-based networks.

Craniocerebral region segmentation Fetal head segmentation methods have been investigated for automatic head-related measurement, which is related to gestational age estimation [12]. In early studies, segmentation is achieved by using classic image processing techniques [13–15] or traditional machine learning methods [16–18]. Until recently, some studies indicate that the craniocerebral regions could be easily segmented by using CNN-based networks. Yaqub [6] designed a CNN classifier that used a transformation from ImageNet to segment the fetal brain regions, with a 0.969 Dice coefficient. Van den Heuvel [19] achieved an accuracy

Fig. 2 Examples of five diseases. The first row shows Blake pouch cyst (**a**), Dandy–Walker malformation (**b**) and cerebellar vermis hypoplasia (**c**) that occur around the cerebellum in the transcerebellar planes. The second row shows ventriculomegaly (**d**) and hydrocephalus (**e**) that occur around the lateral ventricles in the transventricular planes
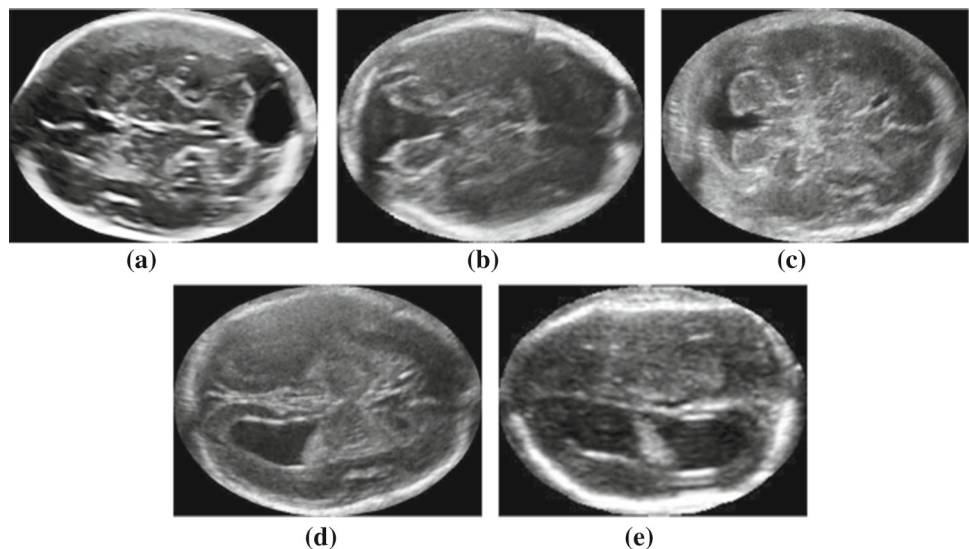
of 0.97 in determining which pixel belonged to the outer edge of a fetal head using U-net-based [20] networks. Based on these studies, we introduced U-net [20] to segment craniocerebral regions, allowing the classification networks to focus on the fetal brain regions.

Weakly supervised localization A visualized interpretation is particularly important in medical tasks. According to [21], a model that can visualize the reasoning process underlying the diagnosis contribute to identifying failure modes [22], building users' trust and assisting doctors in making better decisions [23]. Several studies develop visualization methods by identifying the pixel that has the most impact on the prediction. Exploring the modifications of CNN structures, Zhou [24] proposed class activation mapping (CAM) to obtain class-specific feature maps. Subsequently, Ramprasaath R. [21] introduced a new way of combining feature maps using gradient signals that did not require any modification of the network architecture. For our task, we localized the lesion using CAM for the abnormal cases to provide visualized evidence for the diagnosis.

Methods

Data

We obtained our image dataset from a recently reported original dataset [25], which consisted of a total of 92,748 women with singleton or twin pregnancies with gestational ages between 18 and 32 weeks who underwent prenatal examinations at the 1st Affiliated Hospital of Sun Yat-sen University in China between March 2010 and February 2018. All of the examinations were acquired and diagnosed during routine screenings by a team of 15 doctors with 3–22 years of experience in O&G ultrasound. Ten different ultrasound machines provided by six different manufacturers (GE Voluson 730 Expert/E6/E8/E10, ALOKA SSD-a10, SIEMENS Acuson S2000, TOSHIBA XARIO 200 TUS-X200, SAMSUNG UGEO WS80A and PHILIPS EPIQ7C) were used for data acquisition. The maternal BMI was 24 ± 2.5. The mean gestational age was $22 + 4$ weeks and $26 + 3$ weeks for normal and abnormal cases, respectively. There were three types of images: single-view freezing image saved by the doctors during the exam, split-view freezing image saved during the exam but segmented into individual subimages and video-frame images converted from videos that were exported from 3D volume data [25].

We obtained two separate image datasets from the original dataset for the segmentation and classification task. For segmentation, based on random selection, we obtained 3500, 9850 and 2500 single-view, split-view and video-frame images, respectively. For classification, four categories of fetal brain images were included: normal TV planes, normal TC planes, abnormal TV planes and abnormal TC planes. For normal cases, eligible TV and TC planes followed guidelines set out in the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) [1,4]. For the abnormal cases, TV planes contained occurrences of ventriculomegaly and hydrocephalus and TC planes contained occurrences of Blake pouch cyst (BPC), Dandy–Walker malformations (DWM) and cerebellar vermis hypoplasia (CVH), which are five kinds of common brain malformations. (Examples are shown in Fig. 2.) Additionally, according to the previous study [25], image inclusion and exclusion criteria for the classification were strict and consisted of the following: images without color Doppler or measurement caliper overlays, images where an integrated skull occupied 1/2-2/3 of the screen and images without too many acoustic shad-

Table 1 Details of image counts in the segmentation dataset

	Single-view	Split-view	Video-frame	Total
Training	2000	6000	1500	9500
Validation	750	1850	500	3100
Test	750	2000	500	3250
Total	3500	9850	2500	15,850

ows, where doctors could still recognize the image as one of the SANPs. Note that we obtained all abnormal images of the five diseases. However, in the interest of having balanced data, normal images were randomly selected from the original dataset. As a result, we received 3365, 3595, 2884 and 1801 images of normal TV, normal TC, abnormal TV and abnormal TC planes, respectively. Details are shown in Tables 1 and 2. The study protocol was approved by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-sen University.

Data annotation and division

The ground truth of the craniocerebral region was manually labeled by doctors with an open source ellipse labeling software to construct continuous ellipses along the outer edge of the skull. The dataset was divided into three parts: a training set that was used to train the network, a validation set that was used to select the best model with the smallest loss value and a test set that was used to evaluate the network; all images were randomly assigned to the three sets at the image level with a ratio of approximately 3:1:1. Details are shown in Table 1.

The true label of the images in the classification dataset corresponded to the diagnosis of the cases. The abnormal cases were confirmed by neonatal ultrasound, follow-up examination, or autopsy. The classification dataset was randomly divided into training/validation/test set with a ratio of approximately 3:1:1, on case level rather than image level. Details are shown in Table 2.

The images used to evaluate the lesion localization were randomly selected from the test set of the classification dataset. An expert with 22 years of experience in O&G ultrasound reviewed abnormal images and used an open source rectangle labeling software to construct the bounding box along the lesion area.

Network architectures

Craniocerebral region segmentation

We introduced a fully convolutional network to perform the segmentation task. Specifically, we used the U-net network [20] due to its simple structure and good image processing performance. The architecture of the U-net network is shown in Fig. 3. In the downsampling path, the feature maps were processed by two repeated convolution operations with the ReLU activation function and max-pooling to reduce the resolution. Symmetric structures were obtained in the extension path. After each upsampling operation performed by deconvolution, two repeated convolution operations were performed. The number of channels doubled after max-pooling and halved after deconvolution. The skip connection concatenated the feature maps at the same resolution to reduce the information loss.

In the last layer, a 1×1 convolution was used to map all feature map vectors and to output the prediction mask. The sigmoid function was used to generate a one-channel feature map. We used the Dice loss during training, which is defined as

$$\text{Dice loss} = 1 - \frac{2 \times \sum(y \times y') + \sigma}{\sum y + \sum y' + \sigma} \tag{1}$$

where y and y' are the ground truth and the predicted probability of the image, respectively, and σ is a smoothing factor with the default value of 1. Data augmentation is usually used to prevent overfitting issue [26]. We randomly rotated images within $(-45°, 45°)$ and considered the horizontal or vertical reflections of images. Inspired by [27], we adopted a strategy that used similar quantities of data for the three formats in a batch. Specifically, we used 12 split-view images and 10 of both video-frame and single-view images per batch.

The ellipse annotations were used to generate binary mask images as the ground truth during training. All ultrasound images and the corresponding mask images were converted to grayscale and resized to 256×256. In addition, the input images were normalized to a standard distribution by subtracting the mean value and subsequently dividing by the standard deviation, both of which were calculated for the entire training dataset.

Classification of normal and abnormal brain scans

Deep convolutional neural networks (DCNNs) have powerful feature extraction capabilities [11,26]. Our model shared the basic structure of the VGG-net [11]. Specifically, we had sixteen convolutional layers with a kernel size of 3×3, five max-pooling layers and three FC layers. Moreover, a dropout layer was added after the first two FC layers to reduce overfitting (Table 3).

It has been proven that knowledge transfer is an efficient way of training a high-performance model [28,29]. Despite large differences between natural and ultrasound images, low-level features can be mutually transferred in both domains [7,30,31]. Following the idea in [31], we transferred all parameters from model pretrained on ImageNet to convo-

Table 2 Details of image and case counts (image/case) in the classification dataset

	Normal TV	Normal TC	Abnormal TV	Abnormal TC	Total
Training	2039/2039	2168/2168	1873/216	1082/148	7162/4571
Validation	667/667	717/717	493/74	367/49	2244/1507
Test	659/659	710/710	518/74	352/49	2239/1492
Total	3365/3365	3595/3595	2884/364	1801/246	11,645/7570

Fig. 3 Overview of the classic U-net network architecture. The rectangles denote the feature maps. The numbers next to and above each feature map represent its resolution and the number of channels, respectively. The full network contains a downsampling path, an upsampling path and the skip connection structure
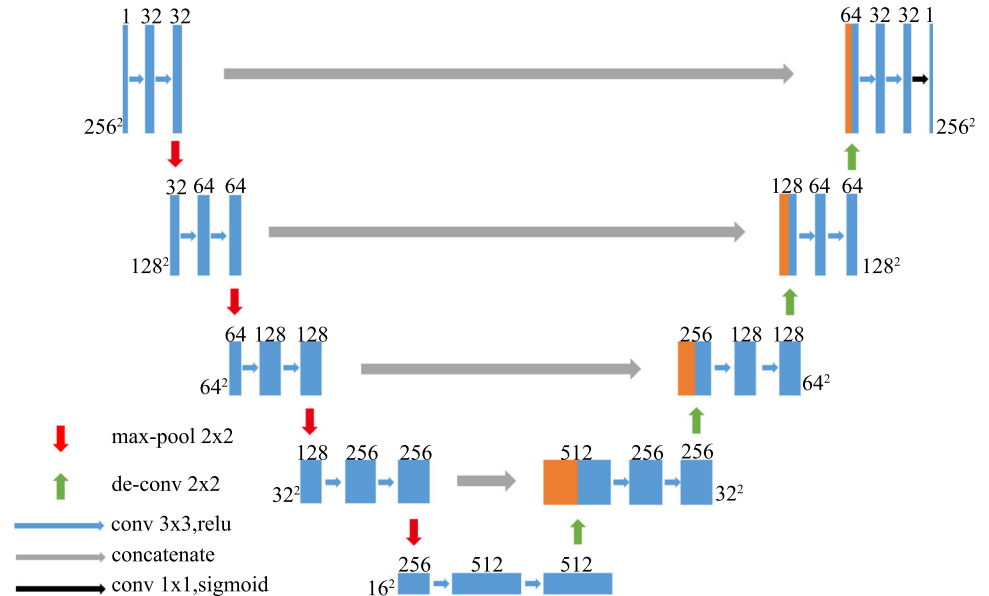

max-pool 2x2

de-conv 2x2

conv 3x3,relu

concatenate

conv 1x1,sigmoid

Table 3 Detailed configuration of our deep convolutional neural network for classification

Input (224 × 224 RGB image)	
Layer	Parameter
Conv. layers 1–2	3×3, 64, stride = 1, padding = 1
Max-pooling	2×2, stride = 2
Conv. layers 3–4	3×3, 128, stride = 1, padding = 1
Max-pooling	2×2, stride = 2
Conv. layers 5–8	3×3, 256, stride = 1, padding = 1
Max-pooling	2×2, stride = 2
Conv. layers 9–12	3×3, 512, stride = 1, padding = 1
Max-pooling	2×2, stride = 2
Conv. layers 13–16	3×3, 512, stride = 1, padding = 1
Max-pooling	2×2, stride = 2
FC1,1×1,1024	
Dropout	
FC2,1×1,1024	
Dropout	
FC3,1×1,4	
Softmax	

lutional layers except the last three FC layers. The FC layers were randomly initialized using Gaussian distribution. Considering the distinct differences between our dataset and the ImageNet dataset, we fine-tuned the pretrained network on our dataset. Specifically, we first fixed all the convolutional layers to train on 10 echoes. Then we adjusted all parameters of model to train on other 20 echoes.

The craniocerebral region images were obtained by a segmentation network. We randomly rotated the images within $(-45°, 45°)$ and considered their horizontal or vertical reflections. Considering the imbalance between various categories, the strategy mentioned in "Craniocerebral region segmentation" section was used. Specifically, we used 8 images for each category per batch. Finally, the three channel input images were resized to 224×224 and normalized to the range $(-1, 1)$ by dividing by 127.5 and subsequently subtracting 1.

The weak supervision localization of lesions

A simple classification result is entirely unconvincing; hence, our model should have the ability to substantiate the diagnosis. In fact, the regions of interest (ROI) corresponding to the identified category can be localized in those feature maps. We believe that the lesions are likely to be associated with the ROI, so the localization of the lesions could be solved by searching the ROI. To identify the regions that significantly affect the final score, we used the gradient-weighted class activation mapping (Grad-CAM) [21] method to produce a
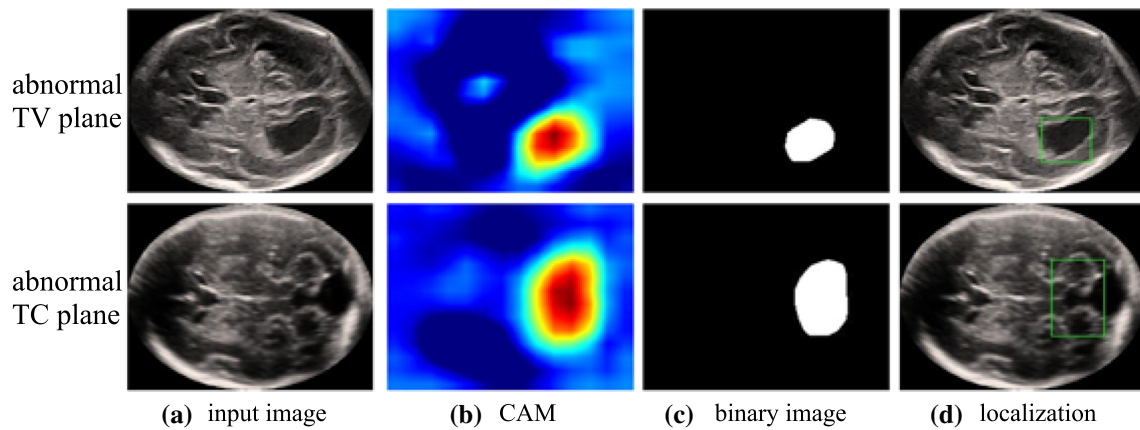
(a) input image   **(b)** CAM   **(c)** binary image   **(d)** localization

Fig. 4 Examples of the localization with bounding boxes in abnormal transventricular (TV) and transcerebellar (TC) planes. **a** Two input images, **b** class activation mapping (CAM) for the two images, **c** the resulting binary images and **d** localization of lesions

Table 4 Segmentation scores for the U-net network

Network	Dice	IOU	Precision	Recall	OR	UR
U-net	0.942	0.909	0.943	0.944	0.047	0.045

CAM and relied on the map to localize the lesions of the input ultrasound images. The core idea of the CAM is to calculate the weight α_k^c that indicates how much the feature map F^k contributes to the final score y^c for a particular class c, which is defined as

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial F_{ij}^k} \tag{2}$$

where F_{ij}^k is the pixel value at location (i, j) of the feature map and Z is the total number of pixels in feature map F^k. Then, we used a weighted combination of the last convolutional feature maps followed by a ReLU function to obtain CAM L_{heatmap}^c corresponding to category c, which is defined as

$$L_{\text{heatmap}}^c = \text{ReLU}\left(\sum_k \alpha_k^c F^k\right) \tag{3}$$

We took the absolute value of the CAM for the abnormal TV and TC planes and converted them to binary values by subjecting them to a threshold of M% of the maximum intensity. The hyperparameter M was empirically set to 25. We considered the largest connected area of the binary image and fitted the minimum rectangular bounding box around it. Some examples are shown in Fig. 4.

Experimental results

Evaluation of the craniocerebral region segmentation

The ground truth masks were generated from the ellipse annotation made by doctors. The predicted craniocerebral region masks for the 3250 test ultrasound images were generated by the U-net network. We compared the predicted masks with the ground truth masks. The results of six common evaluated indices are shown in Table 4. This performance is sufficient for segmentation of the craniocerebral region, and some examples are shown in Fig. 5.

Evaluation of classification for normal and abnormal brain scans

We compared the network predicted labels with the true labels for 2239 test ultrasound images. The precision, recall and F1-score were calculated as shown in Table 5. Additionally, the class confusion matrix obtained with the network is shown in Fig. 6. Both normal and abnormal SANPs are detected with F1-scores above 0.9; thus, most of test images could be correctly classified.

Evaluation of the weak supervision localization of lesions

The ground truth boxes were annotated by experts, and the predicted boxes were generated by fitting the ROI of the network in the abnormal ultrasound images. We compared the network predicted lesion localization boxes with the ground truth boxes for 729 test abnormal ultrasound images. The mean and standard of the intersection over union (IOU) metric were calculated as shown in Table 6. The result demonstrates the ability of the algorithms to localize lesions.
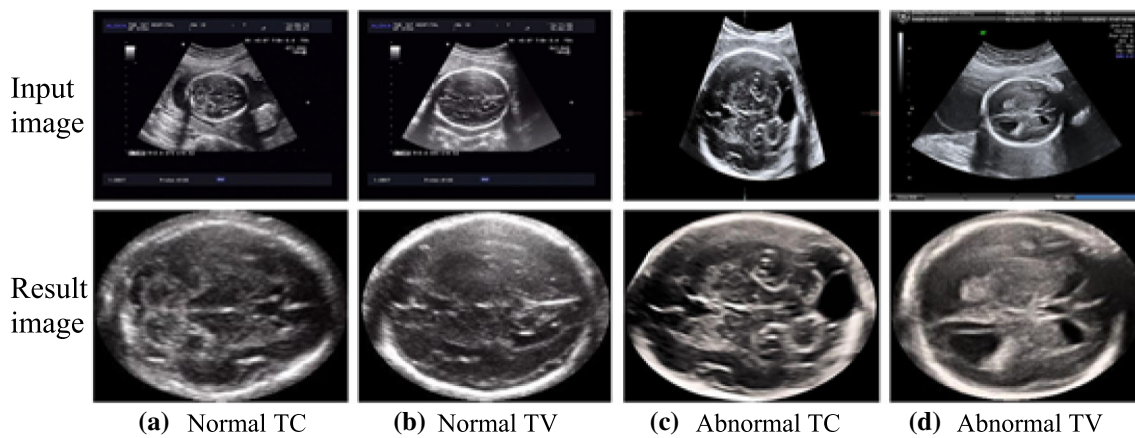
(a) Normal TC   **(b)** Normal TV   **(c)** Abnormal TC   **(d)** Abnormal TV

Fig. 5 Examples of craniocerebral region segmentation obtained by the U-net model

Table 5 Classification scores for the DCNN architecture

Class	Precision	Recall	F1-score
Normal TC	0.978	0.955	0.966
Normal TV	0.989	0.967	0.978
Abnormal TV	0.955	0.988	0.972
Abnormal TC	0.915	0.952	0.934

Table 6 Evaluation scores and counts for five diseases

Disease	Mean IOU	Std. of IOU	No. images
BPC	0.467	0.153	167
CVH	0.458	0.120	90
DWM	0.451	0.135	95
Hydrocephalus	0.566	0.111	200
Ventriculomegaly	0.541	0.114	177
Average	0.497	0.126	N/A
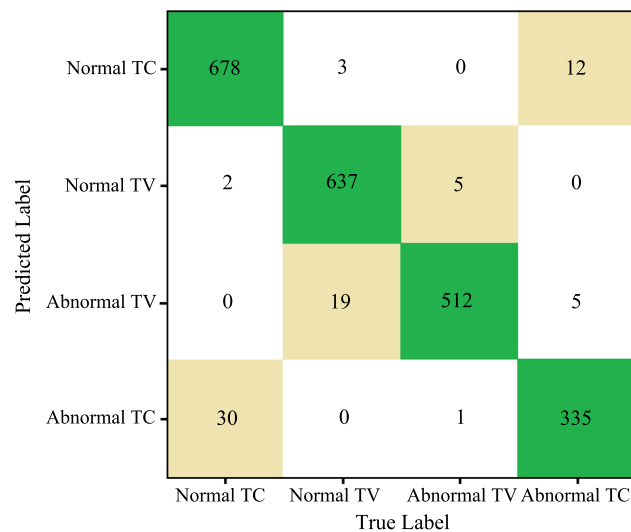
Fig. 6 Class confusion matrix for classification

Discussion

Lower classification accuracies were obtained for the two abnormal planes. As shown in Fig. 8, in the first line, the normal TC planes (the left two images) and the normal TV planes (the right two images) are incorrectly classified as abnormal. We thought that the acoustic shadows appearing in the images could possibly lead to adverse results. If a part of the cerebellum and skull was shaded, the network was likely to classify the images as abnormal.

We observed that the network failed for some abnormal images. This failure may occur for a couple of reasons: the poor quality of the images with more blur and the fewer data with the abnormal planes. Moreover, some anomalies are associated with only subtle findings [1]. As shown in Fig. 8, in the second line, the left two images show the failed abnormal TC planes, and the right two images show the failed abnormal TV planes. The shape of the key structure did not have much deviations from the normal planes.

Note that the network worked on the 2D ultrasound images, data form different machines and types (single-view, split-view and video-frame) were mixed in dataset. The performance of the network is not failed in cases of images from a particular ultrasound machine/type.
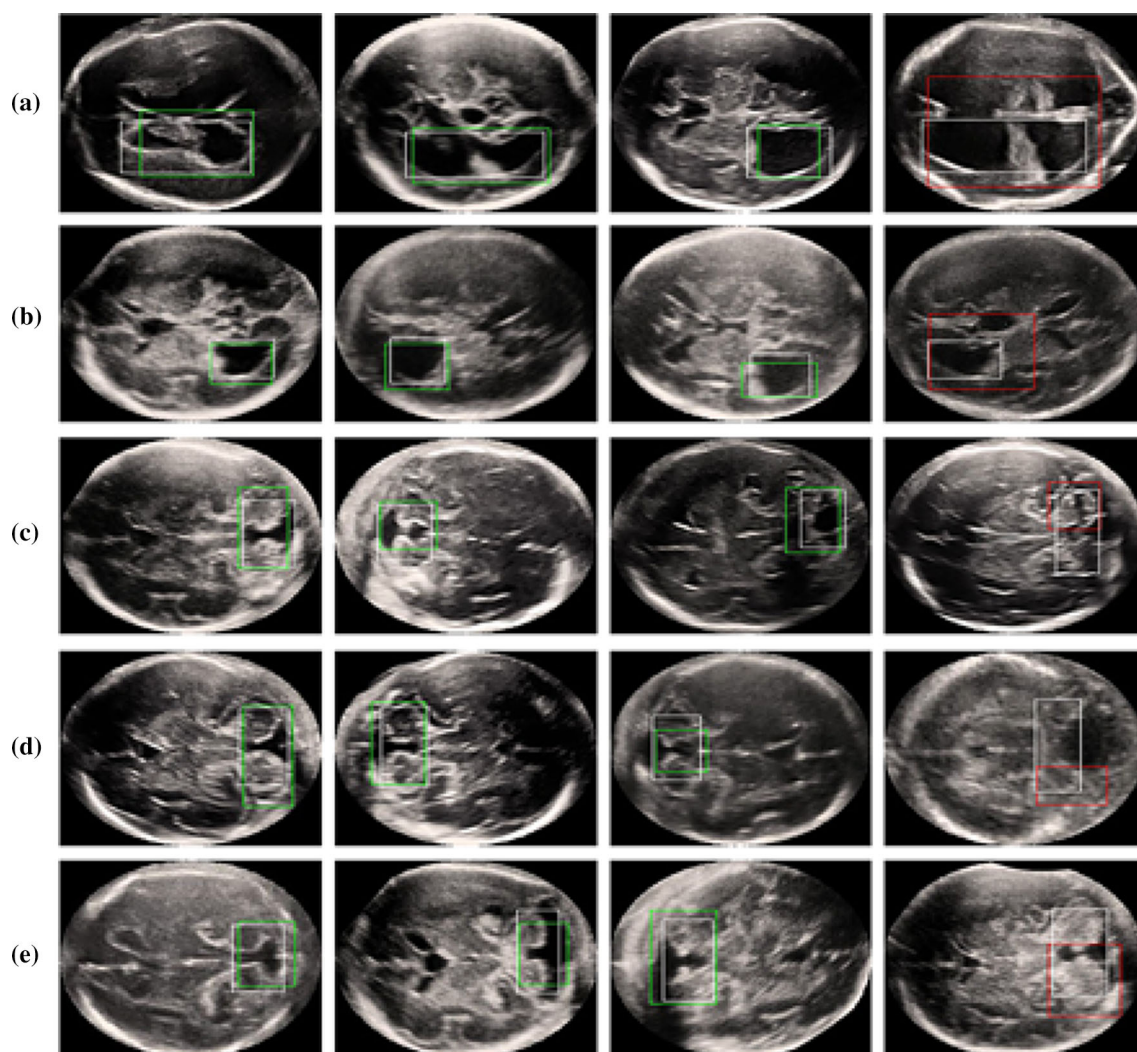
However, improvement is needed to more precisely determine the lesion edges. In Fig. 7, we show examples of the bounding boxes for each of the diseases. These examples prove that our proposed model is able to focus on the lesions that are subject to significant deviations from the normal standard plane. The first three columns illustrate the strong relationship between lesions localization and ground truth with IOU above 0.5, and the last column shows the lower correlation (IOU < 0.5) cases.

Fig. 7 The first three columns for each disease show correct bounding boxes in green (IOU ≥ 0.5), and the last column shows an example of an incorrect box in red (IOU < 0.5). The ground truth bounding boxes are shown in white. **a** Hydrocephalus, **b** ventriculomegaly, **c** BPC, **d** DWM and **e** CVH

In the localization task, some abnormal cases had lower correlations (IOU < 0.5), as shown in Fig. 7, but we observed that the identified localization never completely omitted the lesions. The classifier was expected to learn the symmetric structure of the fetal brain SANPs. In some cases, the localization lost this symmetry and covered only one side of the brain. Some examples were the cases of the posterior fossa anomaly. In other cases, the outlines of the lesions were ambiguous due to severe anomalies, which caused the network to focus on multiple areas and draw larger bounding boxes. Some examples of this issue were cases of hydrocephalus and ventriculomegaly disease.

Limitations

We demonstrated the ability of the algorithms to focus on the lesions and the possibility of obtaining accurate localization. The localization of lesions was currently based purely on the CAM as shown in Fig. 4 and bounding box related to an empirical parameter. Although such method highlighted potential lesions, the IOU of predicted lesion regions by weak supervision was too low. More accurate methods like object detection techniques [32,33] and more principled back-propagated approaches as used in [3] should be considered to improve lesion localization.

Another limitation was the data used in our paper, which were from the same center. More cases from different institutions should be considered for better robustness.
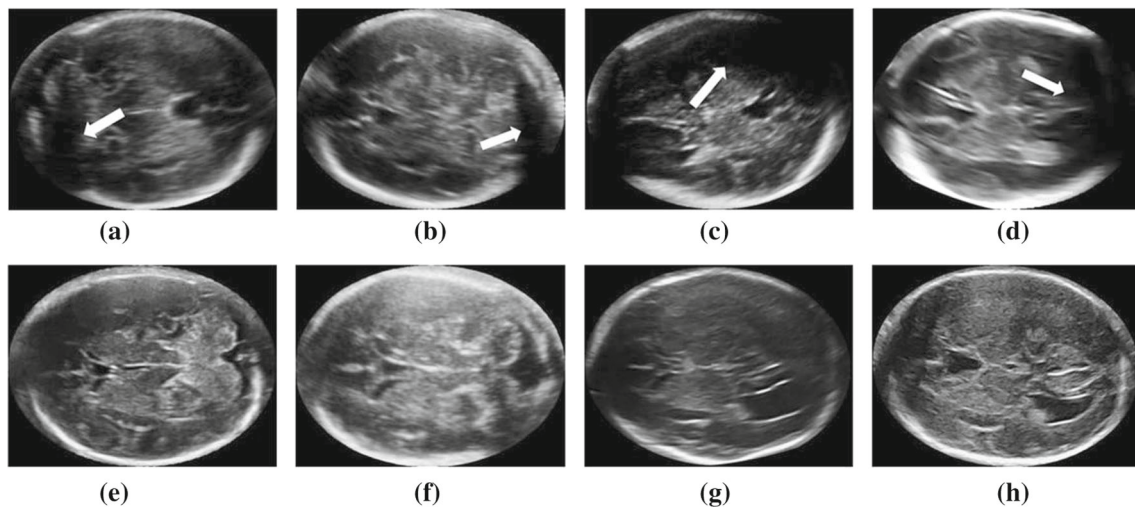
Fig. 8 Examples of incorrectly classified images. The fail normal TC (**a**, **b**) and normal TV (**c**, **d**) planes are incorrectly classified as abnormal due to acoustic shadows (highlighted with arrows). The fail abnormal TC (**e**, **f**) and abnormal TV (**g**, **h**) planes are incorrectly classified as normal due to the difficult detection of subtle malformations

Additionally, the network only diagnosed images of fetal brain transverse standard planes as normal or abnormal for now. In clinical practice, sagittal and coronal planes are also used by doctors in fetal neurosonographic assessment to make a specific diagnosis, e.g., BPC, DWM and ventriculomegaly, rather than normal or abnormal. We need further researches to support more planes.

Last but not least, we removed the images with too much acoustic shadows in this study. However, acoustic shadows are likely lead to adverse results and often appear in real ultrasonic images. Therefore, we shall evaluate the influence in further researches.

Conclusion

In this paper, we developed the first algorithms for fetal brain diseases diagnosis in prenatal ultrasound. Our algorithms exploited U-net [20] to segment the craniocerebral region and VGG-net [11] network to distinguish the normal and abnormal ultrasound images in TV and TC planes. The diagnosis tasks performed well. Additionally, visualized evidence was provided by the localization of the lesions. We believe our algorithms could be potentially applied in diagnosis assistance, and expected to help junior doctors in making clinical decision and reducing false negatives of fetal brain abnormalities.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Retrospective studies The study protocol was approved by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-sen University.

Informed consent For this type of study, formal consent is not required.

Human and animal rights For this type of study, formal consent is not required.

References

1. International Society of Ultrasound in Obstetrics & Gynecology Education Committee (2007) Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. Ultrasound Obstet Gynecol 29(1):109–116. https://doi.org/10.1002/uog.3909
2. Filly RA, Cardoza JD, Goldstein RB, Barkovich AJ (1989) Detection of fetal central nervous system anomalies: a practical level of effort for a routine sonogram. Radiology 172(2):403–408. https://doi.org/10.1148/radiology.172.2.2664864
3. Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, Kainz B, Rueckert D (2017) Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultra-

sound. IEEE Trans Med Imaging 36(11):2204–2215. https://doi.org/10.1109/TMI.2017.2712367

4. Salomon LJ, Alfirevic Z, Berghella V, Bilardo C, Hernandez-Andrade E, Johnsen SL, Kalache K, L K-Y, Malinger G, Munoz H (2011) Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol 37:116–126. https://doi.org/10.1002/uog.8831

5. Yu Z, Tan EL, Ni D, Qin J, Chen S, Li S, Lei B, Wang T (2018) A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. IEEE J Biomed Health Inform 22(3):874–885. https://doi.org/10.1109/JBHI.2017.2705031

6. Yaqub M, Kelly B, Papageorghiou AT, Noble JA (2017) A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. Ultrasound Med Biol 43(12):2925–2933. https://doi.org/10.1016/j.ultrasmedbio.2017.07.013

7. Hao C, Dou Q, Ni D, Cheng JZ, Qin J, Li S, Heng PA (2015) Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: International conference on medical image computing and computer-assisted intervention. https://doi.org/10.1007/978-3-319-24553-9_62

8. Chen H, Wu L, Dou Q, Qin J, Li S, Cheng JZ, Ni D, Heng PA (2017) Ultrasound standard plane detection using a composite neural network framework. IEEE Trans Cybern 47(6):1576–1586. https://doi.org/10.1109/TCYB.2017.2685080

9. Zhang L, Chen S, Chin CT, Wang T, Li S (2012) Intelligent scanning: automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination. Med Phys 39(8):5015–5027. https://doi.org/10.1118/1.4736415

10. Ni D, Yang X, Chen X, Chin CT, Chen S, Heng PA, Li S, Qin J, Wang T (2014) Standard plane localization in ultrasound by radial component model and selective search. Ultrasound Med Biol 40(11):2728–2742. https://doi.org/10.1016/j.ultrasmedbio.2014.06.006

11. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

12. Zhang L, Dudley NJ, Lambrou T, Allinson N, Ye X (2017) Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. J Med Imaging 4(2):024001. https://doi.org/10.1117/1.JMI.4.2.024001

13. Lu W, Tan J, Floyd R (2005) Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform. Ultrasound Med Biol 31(7):929–936. https://doi.org/10.1016/j.ultrasmedbio.2005.04.002

14. Jardim SMGVB, Figueiredo MAT (2005) Segmentation of fetal ultrasound images. Ultrasound Med Biol 31(2):243–250. https://doi.org/10.1016/j.ultrasmedbio.2004.11.003

15. Zhang L, Ye X, Lambrou T, Duan W, Allinson N, Dudley NJ (2016) A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D ultrasound images. Phys Med Biol 61(3):1095–1115. https://doi.org/10.1088/0031-9155/61/3/1095

16. Yaqub M, Kelly B, Papageorghiou AT, Noble JA (2015) Guided random forests for identification of key fetal anatomy and image categorization in ultrasound scans. In: International conference on medical image computing and computer-assisted intervention. https://doi.org/10.1007/978-3-319-24574-4_82

17. Anto EA, Amoah B, Crimi A (2015) Segmentation of ultrasound images of fetal anatomic structures using random forest for low-cost settings. In: Engineering in medicine and biology society. https://doi.org/10.1109/EMBC.2015.7318481

18. Namburete AIL, Noble JA (2013) Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters. In: IEEE International symposium on biomedical imaging. https://doi.org/10.1109/ISBI.2013.6556576

19. van den Heuvel P, Hezkiel S, Stefano dK, Chris L, Bram vG (2019) Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries. Ultrasound Med Biol 45(3):773–785. https://doi.org/10.1016/j.ultrasmedbio.2018.09.015

20. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. https://doi.org/10.1007/978-3-319-24574-4_28

21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: IEEE international conference on computer vision. https://doi.org/10.1109/ICCV.2017.74

22. Hoiem D, Chodpathumwan Y, Dai Q (2012) Diagnosing error in object detectors. In: European conference on computer visions. https://doi.org/10.1007/978-3-642-33712-3_25

23. Johns E, Aodha OM, Brostow GJ (2015) Becoming the expert interactive multi-class machine teaching. In: IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2015.7298877

24. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2015) Learning deep features for discriminative localization. In: IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2016.319

25. Xie H, Wang N, He M, Zhang L, Cai H, Xian J, Lin M, Zheng J, Yang Y (2020) Using deep learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol. https://doi.org/10.1002/uog.21967

26. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Int Conf Neural Inf Process Syst 2:1097–1105. https://doi.org/10.1145/3065386

27. He H, Garcia E (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21(9):1263–1284. https://doi.org/10.1109/tkde.2008.239

28. Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. In: Computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2014.222

29. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? Adv Neural Inf Process Syst 27. arXiv:1411.1792

30. Chen H, Ni D, Qin J, Li S, Yang X, Wang T, Heng PA (2015) Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J Biomed Health Inform 19(5):1627–1636. https://doi.org/10.1109/JBHI.2015.2425041

31. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312. https://doi.org/10.1109/TMI.2016.2535302

32. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

33. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2017.690