IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 10, 2021, accepted July 13, 2021, date of publication July 21, 2021, date of current version August 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3098844

Complementary Network for Accurate Amniotic Fluid Segmentation From Ultrasound Images

SIYU SUN[1], JA-YOUNG KWON[2], YEJIN PARK[2], HYUN CHEOL CHO[1], CHANG MIN HYUN[1], AND JIN KEUN SEO[1], (Member, IEEE)

[1]School of Mathematics and Computing (Computational Science and Engineering), Yonsei University, Seoul 03722, South Korea
[2]Department of Obstetrics and Gynecology, Institute of Women's Life Medical Science, Yonsei University College of Medicine, Seoul 03722, South Korea

Corresponding author: Hyun Cheol Cho (whguscjf55@yonsei.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) under Grant 2015R1A5A1009350 and Grant 2017R1A2B20005661.

ABSTRACT This study presents an automatic method for estimating antenatal amniotic fluid (AF) volume from two-dimensional ultrasound (US) images, which is an important indicator of fetal well-being. This automatic estimation of AF volume (AFV) requires automated segmentation of the AF pocket, which is a challenging task due to its amorphous features and US artifacts, such as reverberation, shadowing, particle noise, and signal dropout. Recently, AF-net, a deep-learning method, has been shown to successfully perform AF pocket segmentation. However, we observed that AF-net is prone to misjudging AF pockets containing severe reverberation artifacts. The proposed method addresses this problem by developing a dual path network, which consists of AF-net as the primary path and an auxiliary network as the secondary path. The auxiliary network is designed to focus on the local area that is likely to be contaminated with the reverberation artifacts. It infers this local region and generates a feature map of the artifacts, incorporating it as prior information into a deep neural network, denoted as RVB-net, for segmenting the reverberation-artifact-contaminated AF region. Finally, the segmentation output from the auxiliary network complements the AF-net. Experimental results show that the proposed dual path network effectively reduces misjudgment of the AF pocket caused by severe reverberation artifacts. The proposed dual path network achieved an average Dice similarity coefficient (DSC) of 0.8599 ± 0.1074 (mean $\pm$ standard deviation) for AF pocket segmentation on the entire evaluation set.

INDEX TERMS Ultrasound image, amniotic fluid, reverberation artifacts, image segmentation, deep learning.

I. INTRODUCTION

Amniotic fluid (AF), the protecting liquid contained in the amniotic cavity, is an essential component for fetal development and maturation during pregnancy [1]. The AF volume (AFV) is an important indicator for reflecting pregnancy progress and fetal development [1]–[3]. Therefore, its assessment is indispensable during an antenatal ultrasound (US) [4], and AFV is commonly estimated by measuring the AF index (AFI) [5].

Manual AFI measurement is not only time-consuming, but also prone to erroneous measurements, as it is highly operator-dependent. Incorrectly measured AFI can contribute to an over- or under-diagnosis of fetal status and inappropriate treatment. Therefore, the fully automatic segmentation of the AF pocket has been in substantial demand. Unfortunately, this task is highly challenging, unlike other ultrasound-based automatic fetal biometric measurements [6]–[9]. The difficulty arises due to amorphous features (i.e., an unspecified variety of shapes and sizes) of the AF pocket and various factors, such as reverberation, the AF mimicking region, floating matter, and incomplete or missing boundary, which lead to limited accuracy of AF pocket segmentation [3], [10].

Recently, a deep learning-based method called AF-net [3] has been shown to successfully deal with these problems and reach the level of clinical application. Although AF-net generally exhibits remarkable performance, it sometimes misjudges some areas of the AF pocket, where there are severe reverberation artifacts, as shown in Fig. 1.

The associate editor coordinating the review of this manuscript and approving it for publication was Junxiu Liu.

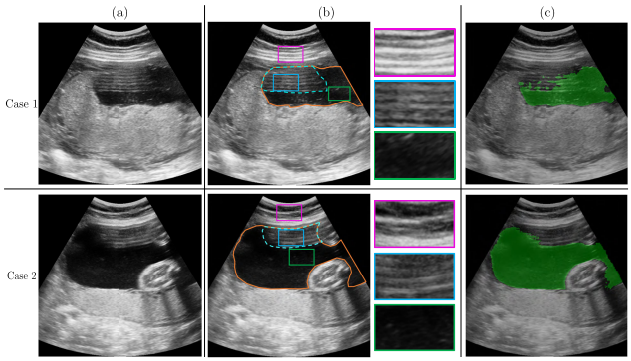


FIGURE 1. Examples of underestimated AF pocket segmentation from AF-net caused by severe reverberation artifacts. (a) Input US images, (b) Ground truths of AF pocket (orange contour) and reverberation artifacts (cyan dashed contour). AF is normally visualized as a dark region with an unspecified variety of shapes and sizes. (c) AF-net segmentation results (green region).

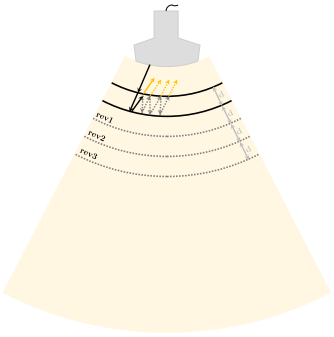


FIGURE 2. Mechanism of reverberation artifacts in US imaging. Reverberation artifacts appear as parallel curves at equal interval (i.e., gray dashed curves annotated by rev1 to rev3) and are produced by the multi-reflection of US beam between two near-field interfaces (black curves).

These severe reverberation artifacts occur when the US beam encounters strong reflectors of the fat-muscle interfaces in the abdominal wall and the transducer-skin interface that are approximately orthogonal to the beam propagation direction. Fig. 2 shows the mechanism of reverberation artifacts in US imaging. Hence, the severe reverberation artifacts in the AF pocket usually have an image pattern of multiple parallel lines at uniform intervals that are typically perpendicular to the direction of US propagation [11]–[14]. This may lead to the misjudgment of the AF region contaminated with the reverberation artifacts as a non-fluid region.

To deal with the severe reverberation artifacts induced misjudgment issue, we proposed a dual path network, whose primary path is AF-net, and the secondary path is an auxiliary segmentation network. See Fig. 3. The auxiliary network is specialized to reduce misjudgment by AF-net. Specifically, it can be regarded as a two-step deep learning approach for segmenting the reverberation artifacts contaminated AF region. The first step infers a local area, where the reverberation artifacts are likely to occur by using the physical correlation among the US, reverberation artifacts, and abdominal wall. In the second step, a feature map of the reverberation artifacts in this local area is first generated and then incorporated into RVB-net, which is developed to segment the severe reverberation artifacts contaminated AF region. This feature map provides supplementary information of the reverberation artifacts into the network. Finally, the segmentation result from the auxiliary network complements the AF-net.

The efficacy of the proposed method has been verified experimentally using a dataset of 2380 clinical 2D US images from 1190 pregnant women. This method performs well on US images affected by severe reverberation artifacts. The auxiliary network, complementary to AF-net, effectively reduces misjudgment of AF pocket caused by the strong reverberation artifacts. Quantitatively, the proposed method yielded an average Dice similarity coefficient (DSC) of 0.8599 ± 0.1074 (mean $\pm$ standard deviation) for AF pocket segmentation, for the entire evaluation set. In addition, an average DSC of 0.9042 ± 0.0361 was obtained, for the US images whose AF pockets were contaminated with severe reverberation artifacts in the evaluation set, i.e., a subset of the evaluation set. For the AF pocket segmentation, the proposed network outperformed other popular deep learning-based segmentation networks, which have shown remarkable overall performance on various medical and natural image segmentation works.

II. METHODS

Let $\boldsymbol{I}$ represent an US image and $\boldsymbol{I}(x, y)$ denote its gray-scale intensity at the pixel position (x, y). The goal is to find an AF pocket segmentation function $f : \boldsymbol{I} \mapsto \mathbb{1}_D$, where $\mathbb{1}_D$ is the indicator function (binary image) of a segmented AF pocket area D. This segmentation function f is to be learned by a labeled training data $\{(\boldsymbol{I}^{(j)}, \mathbb{1}_{D^{(j)}}) : j = 1, \cdots, N\}$. To an approximation, f can be learned by

$$f = \underset{f \in \mathbb{NN}}{\operatorname{argmin}} \frac{1}{N} \sum_{j=1}^{N} \mathcal{L}(f(\boldsymbol{I}^{(j)}), \mathbb{1}_{D^{(j)}}), \tag{1}$$

where $\mathbb{NN}$ is a set of neural networks for segmentation, $\mathcal{L}$ is a loss function for the neural network, and argmin denotes the argument of a minimum.

The proposed deep learning method is a dual path architecture with AF-net as the primary path and an auxiliary semantic segmentation network as the secondary path. The overall process is depicted in Fig. 3. The primary network segments an AF pocket by examining the entire image structure, while the secondary network focuses on a local area, where the primary network may misjudge. The secondary network is designed to reduce the misjudgment of the AF-net. To be precise, the secondary network is specialized in recognizing AF regions that are contaminated with severe reverberation artifacts as follows: first, the secondary path segments an abdominal wall via U-net, determines a local area where the reverberation artifacts are likely to occur, and then generates a derivative feature map of the determined local area. The RVB-net segments a reverberation-artifact-contaminated AF

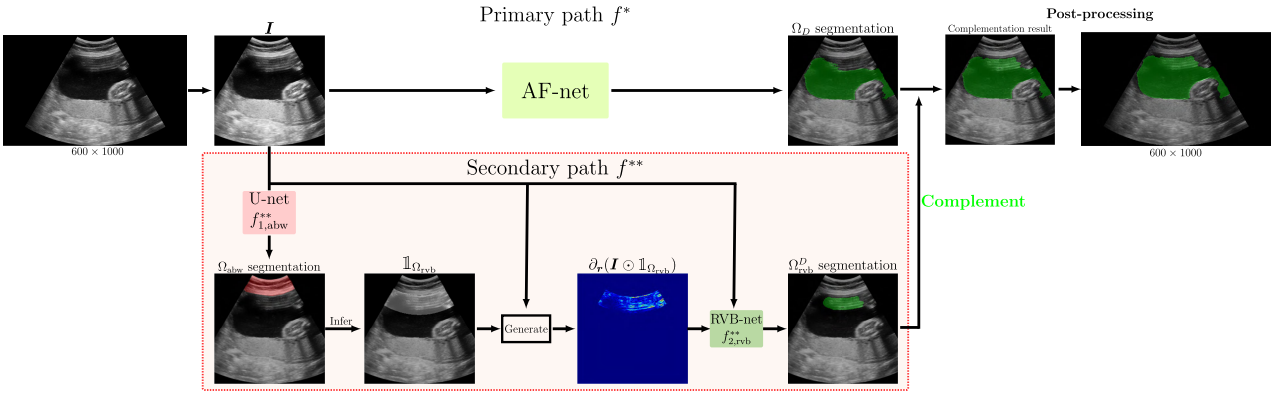


FIGURE 3. Overall structure of the proposed dual path network for accurate AF pocket segmentation. Although AF-net (primary path) alone provides high performance of AF pocket segmentation, we use the secondary path for further improvement. The secondary network is designed to focus on the area that is likely to be contaminated with reverberation artifacts, where the primary path is prone to misjudgment when identifying AF pocket.

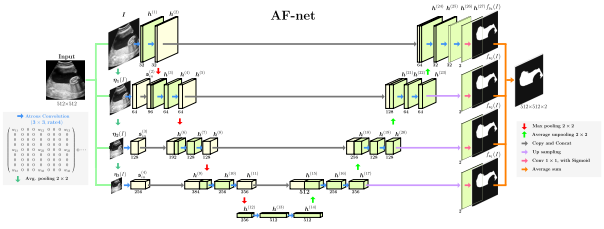


FIGURE 4. Architecture of AF-net for AF pocket segmentation from US image.

region by using the feature map and US image $\boldsymbol{I}$ as an input. A final output is obtained by taking union of the primary and secondary path results, post-processing (morphological opening and closing), and rescaling. The details of each process would be described in the following subsections.

A. PRIMARY PATH: AF-NET

The primary path is AF-net, which is the optimal combination of three techniques; atrous convolution [15], [16], a multi-scale side-input layer, and a side-output layer [17]. The architecture of AF-net is shown in Fig. 4. In AF-net, the segmentation network f^* is learned by

$$f^* = \mathrm{argmin} \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{4} \mathcal{L}_{\mathrm{cross}}(f_{\mathfrak{s}_k}(\boldsymbol{I}^{(j)}), \mathbb{1}_{D^{(j)}}), \tag{2}$$

where $\mathcal{L}_{\mathrm{cross}}(f_{\mathfrak{s}_k}(\boldsymbol{I}), \mathbb{1}_D)$ is the cross entropy loss function given by

$$\mathcal{L}_{\mathrm{cross}}(f_{\mathfrak{s}_k}(\boldsymbol{I}), \mathbb{1}_D) = -\sum_{(x,y)\in\Omega} \big[\mathbb{1}_D(x,y) \log f_{\mathfrak{s}_k}(\boldsymbol{I})(x,y) + (1-\mathbb{1}_D(x,y)) \log(1-f_{\mathfrak{s}_k}(\boldsymbol{I})(x,y))\big]. \tag{3}$$

Here, Ω is the area of the domain of input image $\boldsymbol{I}$, and $f_{\mathfrak{s}_k}(\boldsymbol{I})$ is the k-th side output of AF-net. As shown in Fig. 4, the four side-outputs (i.e., $f_{\mathfrak{s}_k}(\boldsymbol{I})$, $k = 1, \ldots, 4$) are produced from early layers in the decoding path. For example, the side output $f_{\mathfrak{s}_1}(\boldsymbol{I})$ is computed as follows:

$$f_{\mathfrak{s}_1}(\boldsymbol{I}) = \sigma(\mathit{Ave\text{-}unpool}(\boldsymbol{h}^{(17)}, 8) \circledast \boldsymbol{W}^{\mathfrak{s}_1}), \tag{4}$$

where σ is the sigmoid function, $\mathit{Ave\text{-}unpool}(\boldsymbol{h}^{(17)}, 8)$ is the bilinear upsampling of a factor eight to the feature map $\boldsymbol{h}^{(17)}$, $\circledast$ is standard convolution, and $\boldsymbol{W}^{\mathfrak{s}_1}$ is a set of 1×1 filters. These side-outputs help train intermediate layers by backpropagating the side-output loss to early layers in the decoding path [3].

In the encoding path, the first feature map $\boldsymbol{h}^{(1)}$ is produced by $\boldsymbol{h}^{(1)} = \mathit{ReLU}(\boldsymbol{I} \circledast^4 \boldsymbol{W}^{(1)})$, where $\circledast^4$ stands for atrous convolution with rate 4, $\boldsymbol{W}^{(1)}$ is a set of 3×3 filters, and $\mathit{ReLU}(x) = \mathit{max}\{x, 0\}$ is the rectified linear unit. Similarly, $\boldsymbol{h}^{(2)} = \mathit{ReLU}(\boldsymbol{h}^{(1)} \circledast^4 \boldsymbol{W}^{(2)})$. Atrous convolution plays a key role in handling the confusing factors (i.e., reverberation artifacts, AF mimicking region, floating matter, and missing or incomplete boundary) by enlarging the receptive field size of the network. For $\boldsymbol{h}^{(3)}$, we use the side input $\eta_1(\boldsymbol{I}) = \mathit{Ave\text{-}pool}(\boldsymbol{I}, 2)$, which is the average-pooling image of $\boldsymbol{I}$ by a factor of two. More precisely, $\boldsymbol{h}^{(3)}$ is obtained by concatenation of $\boldsymbol{h}^{(2)}$ and $\mathit{ReLU}(\eta_1(\boldsymbol{I}) \circledast^4 \boldsymbol{W}^{(3)})$. Similarly, $\boldsymbol{h}^{(6)}$ and $\boldsymbol{h}^{(9)}$ are produced by using $\eta_2(\boldsymbol{I})$ and $\eta_3(\boldsymbol{I})$ that are the average pooling of $\boldsymbol{I}$ by a factor of four and eight, respectively. This multi-scale side-input layer achieves multiple levels of receptive field size in the encoding path of AF-net, therefore it helps deal with unpredictable shapes and large size variations of AF pockets.

The combination of these three techniques produces synergistic effect, which effectively deals with various factors and amorphous features of the AF pocket. For a detailed explanation of AF-net, refer to [3].

B. SECONDARY PATH: AUXILIARY NETWORK TO REDUCE MISJUDGMENT OF AF-NET

The secondary path is an auxiliary network that serves to correct reverberation artifact-induced misjudgment of AF-net.

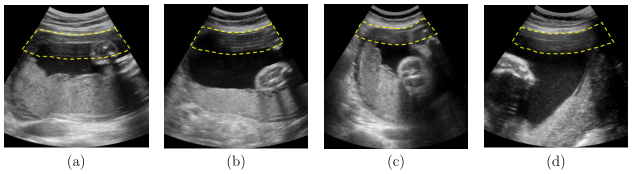


FIGURE 5. The attention area Ω_{rvb} (inside the yellow contour) that is likely to be contaminated by reverberation artifacts.

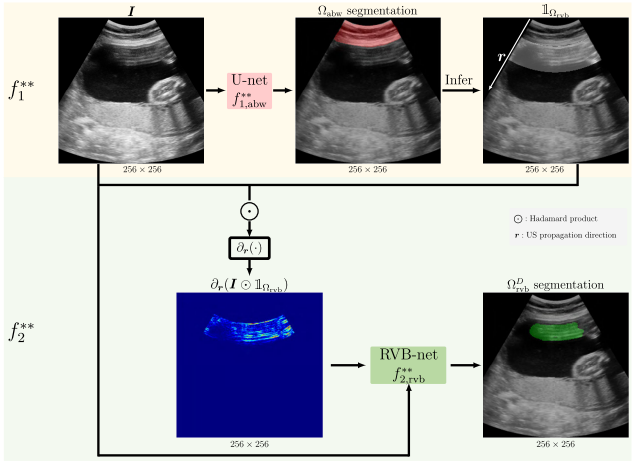


FIGURE 6. Schematic diagram of the secondary path f^{**} consisting of $f_1^{**}: I \mapsto \mathbb{1}_{\Omega_{\text{rvb}}}$ and $f_2^{**}: (I, \mathbb{1}_{\Omega_{\text{rvb}}}) \mapsto \mathbb{1}_{\Omega_{\text{rvb}}^D}$. The first network f_1^{**} infers Ω_{rvb}. The second network f_2^{**} generates derivative feature map of inferred area Ω_{rvb} and then segments Ω_{rvb}^D by using derivative feature as prior information of reverberation artifacts.

It focuses on the region of interest, denoted by Ω_{rvb}, that is likely to be contaminated by reverberation artifacts in image $\boldsymbol{I}$. Fig. 5 shows the region of Ω_{rvb}. The secondary path f^{**} comprises two steps: the first step is to learn function $f_1^{**}: \boldsymbol{I} \mapsto \mathbb{1}_{\Omega_{\text{rvb}}}$ for determining Ω_{rvb} from $\boldsymbol{I}$. The second step is to learn $f_2^{**}: (\boldsymbol{I}, \mathbb{1}_{\Omega_{\text{rvb}}}) \mapsto \mathbb{1}_{\Omega_{\text{rvb}}^D}$ for segmenting Ω_{rvb}^D, where Ω_{rvb}^D is the reverberation artifacts contaminated area inside D. Fig. 6 shows the overall structure of the secondary path.

1) FIRST STEP: DETERMINING Ω_{rvb}

The key concept behind the determination of Ω_{rvb} is to use the physical correlation among the US, reverberation artifacts and abdominal wall. The reverberation artifacts usually occur beneath the abdominal wall when the US beam meets the fat-muscle interfaces in the abdominal wall, and it is also observed that the thickness of these artifacts is almost equal to the thickness of the abdominal wall. Let Ω_{abw} denote the abdominal wall occupying the area between the skin and a thick fascial layer named the peritoneum-fascia layer [18]. Then, Ω_{rvb} is inferred as the reflection area of Ω_{abw} on the fascial layer along the US beam propagation direction, as illustrated in Fig. 7.

The first step f_1^{**} is to segment Ω_{abw}, which is used to infer Ω_{rvb}, as illustrated in Fig. 6. U-net [19] is adopted to

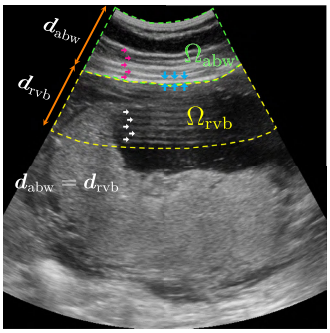


FIGURE 7. Inference of the attention area Ω_{rvb} by using the physical correlation among US, reverberation artifacts and abdominal wall. Ω_{abw} and Ω_{rvb} are inside the green and yellow contours, respectively. The magenta, white and blue arrows point to the fascia, reverberation artifacts in AF pocket and the peritoneum-fascia layer, respectively.

TABLE 1. Architecture of U-net for segmenting Ω_{abw}.

	Layer	Input	Output	Kernel	Activation
Encoder	conv(× 2) 1	256 × 256 × 1	256 × 256 × 32	3^2 × 32	ReLU
	max_pool 1	256 × 256 × 32	128 × 128 × 32	2^2	—
	conv(× 2) 2	128 × 128 × 32	128 × 128 × 64	3^2 × 64	ReLU
	max_pool 2	128 × 128 × 64	64 × 64 × 64	2^2	—
	conv(× 2) 3	64 × 64 × 64	64 × 64 × 128	3^2 × 128	ReLU
	max_pool 3	64 × 64 × 128	32 × 32 × 128	2^2	—
	conv(× 2) 4	32 × 32 × 128	32 × 32 × 256	3^2 × 256	ReLU
	max_pool 4	32 × 32 × 256	16 × 16 × 256	2^2	—
	conv(× 2) 5	16 × 16 × 256	16 × 16 × 512	3^2 × 512	ReLU
Decoder	Upsample & Skip connection 1	16 × 16 × 512	32 × 32 × 512	2^2	—
	Conv(× 2) 6	32 × 32 × 512	32 × 32 × 256	2^2 × 256	ReLU
	Upsample & Skip connection 2	32 × 32 × 256	64 × 64 × 256	2^2	—
	Conv(× 2) 7	32 × 32 × 512	32 × 32 × 256	2^2 × 256	ReLU
	Upsample & Skip connection 3	64 × 64 × 128	128 × 128 × 128	2^2	—
	Conv(× 2) 8	128 × 128 × 128	128 × 128 × 64	2^2 × 64	ReLU
	Upsample & Skip connection 4	128 × 128 × 64	256 × 256 × 64	2^2	—
	Conv(× 2) 9	256 × 256 × 64	256 × 256 × 32	2^2 × 32	ReLU
	Conv & Sigmoid	256 × 256 × 32	256 × 256 × 2	1^2 × 2	—

learn the abdominal wall segmentation function $f_{1,\text{abw}}^{**}: \boldsymbol{I} \mapsto \mathbb{1}_{\Omega_{\text{abw}}}$. The network, described in Table 1, allows to achieve an accurate segmentation of the abdominal wall. This is because abdominal wall has a distinct morphological feature. We use a labeled training data $\{(\boldsymbol{I}^{(j)}, \mathbb{1}_{\Omega_{\text{abw}}^{(j)}}) \mid j = 1, \cdots, N\}$ to learn $f_{1,\text{abw}}^{**}$ by

$$f_{1,\text{abw}}^{**} = \text{argmin} \frac{1}{N} \sum_{j=1}^{N} \mathcal{L}_{\text{cross}}(f_{1,\text{abw}}^{**}(\boldsymbol{I}^{(j)}), \mathbb{1}_{\Omega_{\text{abw}}^{(j)}}), \quad (5)$$

where $\mathcal{L}_{\text{cross}}$ is the cross entropy loss described in Eq. (3).

2) SECOND STEP: SEGMENTATION OF Ω_{rvb}^D

To allow the network to pay more attention to the distinct pattern of the severe reverberation artifacts in the AF pocket, the second step first generates a derivative feature map $\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})$, which is the derivative of $\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}$ along US beam propagation direction $\boldsymbol{r}$. Here, $\odot$ denotes the Hadamard product. This generation is motivated by the observation that the reverberation artifacts are most intense when US beam meets the fat-muscle interfaces in an orthogonal direction (Fig. 6).

The feature map $\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})$, as prior information of reverberation artifacts in the inferred region Ω_{rvb}, is then

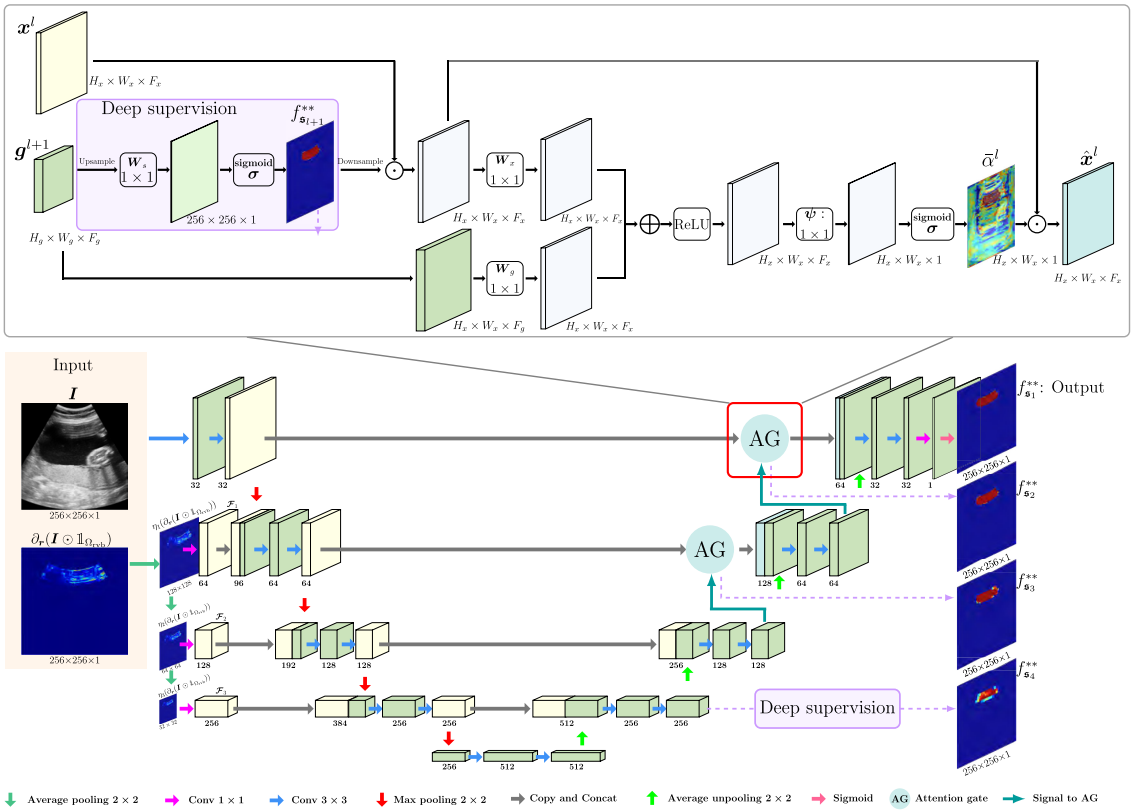


FIGURE 8. Architecture of RVB-net for Ω^{D}_{rvb} segmentation: the input is $(\boldsymbol{I}, \partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}))$, and the output $f^{**}_{\mathfrak{s}_1}$ is a probability map of Ω^{D}_{rvb} on the last layer. The generated feature map $\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})$ is incorporated into the network as side inputs, which introduce multi-scale features of reverberation artifacts. The AG makes the network focus on the local area. These two modules boost the segmentation performance of Ω^{D}_{rvb}.

incorporated into a semantic segmentation network called the RVB-net to segment Ω^{D}_{rvb}. This is to learn a function $f^{**}_{2,\text{rvb}} : \boldsymbol{I}_{\text{rvb}} \mapsto \mathbb{1}_{\Omega^{D}_{\text{rvb}}}$, where $\boldsymbol{I}_{\text{rvb}} := (\boldsymbol{I}, \partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}))$ indicates the input. RVB-net uses labeled training data $\{(\boldsymbol{I}^{(j)}_{\text{rvb}}, \mathbb{1}_{\Omega^{D(j)}_{\text{rvb}}}) \,|\, j{=}1, 2, \cdots, N\}$ to learn $f^{**}_{2,\text{rvb}}$, which is achieved by

$$f^{**}_{2,\text{rvb}} = \text{argmin} \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{4} \mathcal{L}_{\text{cross}}(f^{**}_{\mathfrak{s}_k}(\boldsymbol{I}^{(j)}_{\text{rvb}}), \mathbb{1}_{\Omega^{D(j)}_{\text{rvb}}}), \quad (6)$$

where $\mathcal{L}_{\text{cross}}$ is the cross entropy loss described in Eq. (3), and $f^{**}_{\mathfrak{s}_k}$ is the k-th (k=1, 2, 3, 4) side output.

The RVB-net, depicted in Fig. 8, is a deeply supervised network, equipped with two important modules: a side-input layer and an attention gate (AG).

a: REVERBERATION-ENRICHED SIDE INPUT

The side input of RVB-net is a pyramid of multi-scale derivative feature maps, which are used as supplementary information input into the network consecutively. As shown in Fig. 8, the i-th side input ($i = 1, 2, 3$) is given by

$$\eta_i(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})) = \textit{Ave-pool}(\eta_{i-1}(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})), 2), \quad (7)$$

where $\eta_0(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}) = \partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})$. Each side input $\eta_i(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}))$ generates a feature map $\boldsymbol{\mathcal{F}}_i$ through a 1×1 standard convolution, given by

$$\boldsymbol{\mathcal{F}}_i = ReLU(\eta_i(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}})) \circledast \boldsymbol{W}^{\boldsymbol{\mathcal{F}}_i}), \quad i \in \{1, 2, 3\}, \quad (8)$$

where $W^{\boldsymbol{\mathcal{F}}_i}$ is a set of 1×1 filters. Then, $\boldsymbol{\mathcal{F}}_i$ is concatenated to the encoding path, as shown in Fig. 8. Our side input module has the following advantages:

- The use of $\{\eta_i(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}))\}_{i=1}^{3}$ allows the network to focus more on the local region Ω_{rvb} and facilitates the network to learn the features of the reverberation artifacts.
- For the image whose AF pocket is contaminated with the reverberation artifacts, the multi-scale feature map $\{\eta_i(\partial_{\boldsymbol{r}}(\boldsymbol{I} \odot \mathbb{1}_{\Omega_{\text{rvb}}}))\}_{i=1}^{3}$ allows the network to learn multi-scale contextual features of the reverberation artifacts, and thus it helps improve the accuracy of segmenting Ω^{D}_{rvb}.

b: ATTENTION GATE

Our proposed attention gate is designed to guide the network to focus on the target region Ω^{D}_{rvb}. It is performed before the skip connection to suppress irrelevant regions and highlight the features around the target region.

Let $\boldsymbol{x}^l = \{\boldsymbol{x}_i^l\}_{i=1}^n$ denote the low-level features in the encoding path, where each $\boldsymbol{x}_i^l \in \mathbb{R}^{F_x}$ is a pixel-wise feature vector of length F_x. The global features, denoted as $\boldsymbol{g}^{l+1}$, are features at a coarser scale in the decoding path. AG yields attention coefficient $\alpha^l = \{\alpha_i^l\}_{i=1}^n$, where $\alpha_i^l \in [0, 1]$, that prunes $\boldsymbol{x}^l$ to output $\hat{\boldsymbol{x}}^l = \{\alpha_i^l \boldsymbol{x}_i^l\}_{i=1}^n$. The attention coefficient α_i^l prunes $\boldsymbol{x}_i^l$ by the element-wise product and preserves feature activations only relevant to Ω_{rvb}^D segmentation. Specifically, a side probability map of Ω_{rvb}^D, denoted as $f_{\mathfrak{s}_{l+1}}^{**}$, is first generated from $\boldsymbol{g}^{l+1}$:

$$f_{\mathfrak{s}_{l+1}}^{**} = \sigma(\textit{Ave-unpool}(\boldsymbol{g}^{l+1}, 2^l) \circledast \boldsymbol{W}_s), \tag{9}$$

where $\boldsymbol{W}_s \in \mathbb{R}^{F_g \times 1}$ is a set of 1×1 filters, and $l = 1, 2$. Each side probability map $f_{\mathfrak{s}_{l+1}}^{**}$ is supervised by minimizing the loss defined in Eq. (6). Here, for notational convenience, we denote $f_{\mathfrak{s}_{l+1}}^{**}$ by s^{l+1}. Then, s^{l+1} is used to scale $\boldsymbol{x}^l$ by $\{s_i^{l+1}\boldsymbol{x}_i^l\}_{i=1}^n$. Afterwards, $\{s_i^{l+1}\boldsymbol{x}_i^l\}_{i=1}^n$ and $\boldsymbol{g}^{l+1}$ are used jointly to produce a sub-attention coefficient $\bar{\alpha}$ as:

$$\begin{aligned} q_{att,i}^l &= \boldsymbol{\psi}^T(ReLU(\boldsymbol{W}_x^T(s_i^{l+1}\boldsymbol{x}_i^l) + \boldsymbol{W}_g^T\boldsymbol{g}_i^{l+1} + \boldsymbol{b}_g)) + b_\psi, \\ \bar{\alpha}^l &= \sigma(q_{att}^l(\boldsymbol{x}^l, \boldsymbol{g}^{l+1}; \boldsymbol{\Theta}_{att})), \end{aligned} \tag{10}$$

where $\boldsymbol{\Theta}_{att} = [\boldsymbol{W}_x \in \mathbb{R}^{F_x \times F_x}, \boldsymbol{W}_g \in \mathbb{R}^{F_g \times F_x}, \boldsymbol{\psi} \in \mathbb{R}^{F_x \times 1}, \boldsymbol{b}_g \in \mathbb{R}^{F_x}, b_\psi \in \mathbb{R}]$ is a parameter set consisting of 1×1 filters and bias terms. Hence, the attention coefficient α^l is formulated as

$$\alpha^l = \bar{\alpha}^l \odot s^{l+1}. \tag{11}$$

In $\{s_i^{l+1}\boldsymbol{x}_i^l\}_{i=1}^n$, the irrelevant features in the background are gradually filtered out by the saliency probability map s^{l+1}. Using this scaled feature instead of $\boldsymbol{x}^l$ to generate the attention coefficient, the model activations can focus on the target region in Ω_{rvb} and the experimental results show that the AG improves the segmentation performance.

C. COMPLEMENTATION AND POST-PROCESSING

The segmented Ω_{rvb}^D from the secondary path complements the AF segmentation from the primary path. The complemented result $f(\boldsymbol{I})$ is represented as

$$f(\boldsymbol{I})(x, y) = \begin{cases} 1 & \text{if } f^*(\boldsymbol{I})(x, y) = 1 \text{ or } f^{**}(\boldsymbol{I})(x, y) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $f^{**}(\boldsymbol{I})$ is resized to the same size of $f^*(\boldsymbol{I})$ before complementation. Then, post-processing is performed to fine-tune the complementation result, including morphological operation [20], [21] and rescaling. Specifically, the opening is used to reduce the small false positive errors, while closing is to fill small holes. Here, a 10×10 square structuring element is utilized. Finally, the segmentation map is rescaled into the original 600×1000 size. An illustration of this procedure is shown in Fig. 9.

III. EXPERIMENTS AND RESULTS

A. DATASET

For training and evaluation, a dataset of total 2380 US images of AF pocket from 1190 pregnant women who underwent antenatal scanning at the Maternal-Fetal Division, Department of Obstetrics and Gynecology, Yonsei University College of Medicine, Seoul, Korea were used (IRB no.: 1-2019-0052). Images were acquired by expert sonographers with W10 and WS80A US machines (SAMSUNG Medison, Seoul, Korea) using a 2-6 MHz transabdominal transducer CA1-7A. The gestation is between 20 + 0 and 36 + 6 weeks of pregnant women who are over 19 years old and are Korean. The pregnant women with abnormally reduced amniotic fluid, such as anhydramnios/oligohydramnios due to abnormal urine, and premature preterm rupture of membrane are excluded. Various conditions of pregnant women that may affect the quality of US images have been considered, including the subcutaneous fat distance, the position and size of AF pocket, and fat distribution.

Among 2380 US images, only 90 images (3.8% of the total) have reverberation artifacts in AF pocket. Hence, while training the secondary path, a data imbalance problem arises. To balance the data distribution, we choose 360 images (60 images w/ RVB and 300 images w/o RVB) as the training data of the secondary path. To train the primary path, 2200 images (60 images w/ RVB and 2140 images w/o RVB) were used. The remaining 180 images (30 images w/ RVB

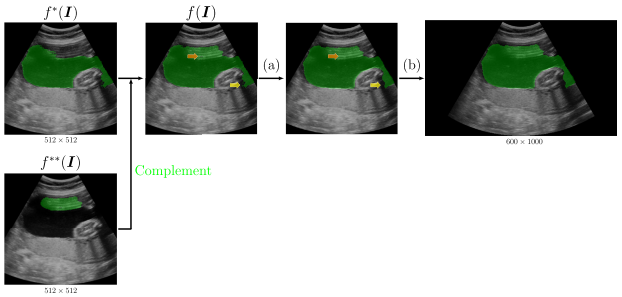


FIGURE 9. The illustration of complementation and post-processing. The complemented result $f(\boldsymbol{I})$ may contain a few small holes (indicated by orange arrows) or false positive errors (indicated by yellow arrows). Hence, morphological opening and closing (denoted as (a)) are performed to fine tune it. Finally, the result is rescaled to original image size, denoted as (b).

TABLE 2. Dataset summary. In the experimental section, "w/", "w/o", and "RVB" denote an abbreviation for with, without and reverberation artifacts, respectively. (a) Data splitting for training and evaluation of the proposed method. (b) Training data of the primary and secondary paths, respectively.

(a)

	Number ofi mages			Number of subjects		
	w/ RVB	w/o RVB	Total	w/ RVB	w/o RVB	Total
Training set	60	2140	2200	40	1000	1040
Evaluation set	30	150	180	30	120	150
Total	90	2290	2380	70	1120	1190

(b)

		Number ofi mages			Number of subjects		
		w/ RVB	w/o RVB	Total	w/ RVB	w/o RVB	Total
Training set	Primary path (AF-net)	60	2140	2200	40	1000	1040
	Secondary path		300[a]	360		150	190

[a] The 300 data w/o RVB were selected randomly from 2140 data w/o RVB, which are in the training data of the primary path.

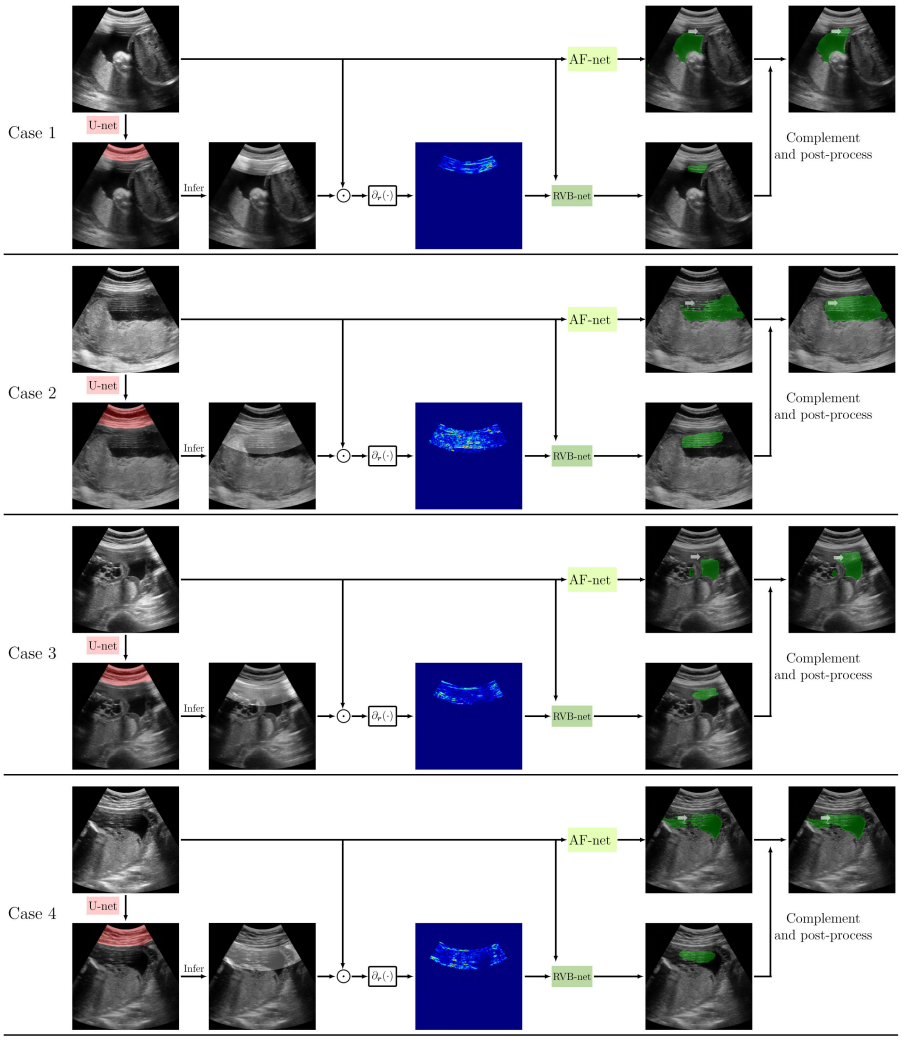


FIGURE 10. Examples of complementary results of dual path network. From Case 1 to Case 4, the AF pocket is contaminated with a strong reverberation artifacts (indicated by a white arrow).

and 150 images w/o RVB) were used for evaluation. The details for the same can be found in the Table 2. For a fair comparison, other popular segmentation networks (i.e., U-net, Attention U-net, Deeplabv3+, and SegNet) were trained by 2200 images and evaluated by the same 180 images. All ground truths are delineated by two sonographers and then verified by experts to ensure the quality of the delineations.

B. DATA PREPROCESSING

Our preprocessing involves cropping, resizing, normalization, and augmentation techniques. The original 600 × 1000 images were first cropped into images of 600 × 600 pixels, which aims to remove the unnecessary two sides of the image and make training efficient. The bilinear interpolation was then used to resize images to 512×512, and the image intensities were normalized to [0,1], which were used as inputs of the primary path. To make the secondary path computationally efficient, the images were resized to 256 × 256 pixels by using bilinear interpolation. To build a robust reverberation artifacts segmentation model, data augmentation techniques were adopted, including random brightness settings, contrast, and horizontal flipping. The data augmentation technique was adopted in the training process of each neural network. The brightness of the image was first adjusted by adding a randomly picked float number in the interval (-0.125,0.125) to all pixels of the images. The image contrasts were then adjusted with a contrast factor that was randomly picked within (0.5,1.5) [22]. The images were then randomly flipped in a horizontal manner. In addition, these augmentation techniques were performed with a probability of 0.5 (i.e., there is a 0.5 chance of each augmentation).

C. IMPLEMENTATION DETAILS

The proposed dual path network is implemented with Python 3.7.1 and Tensorflow r1.13 [23], [24]. During training, Adam stochastic optimization is employed to optimize each network with batch normalization, a mini-batch of four images, and an initial learning rate of 10^{-4} with a ''poly'' learning

TABLE 3. Segmentation evaluation of dual path network on evaluation data. (a) Comparison of AF pocket segmentation between AF-net and dual path network ("n.s." stands for not statistically significant), (b) abdominal wall segmentation for all evaluation data, (c) segmentation result of reverberation artifacts contaminated AF region from RVB-net.

Data	Network	DSC		Sensitivity		Precision		HD		ASSD		RMSD	
		Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value
w/ RVB	AF-net	0.8795± 0.0508	50e-04	0.8250± 0.0921	12e-05	0.9495± 0.0390	20e-02	149110± 9.9218	n.s.	2.3469± 1.1443	14e-03	3.7174± 1.9065	47e-03
	Dual path	0.9042 ± 0.0361	—	0.8863 ± 0.0620	—	0.9287± 0.0612	—	14.4083± 9.9300	—	1.9335 ± 0.8870	—	3.2230 ± 1.6716	—
w/o RVB	AF-net	0.8510± 0.1165	n.s.	0.8030± 0.1517	n.s.	0.9302± 0.0717	n.s.	15.6175± 14.2120	n.s.	3.0810± 8.6951	n.s.	4.6652± 9.0421	n.s.
	Dual path	0.8510± 0.1165	—	0.8030± 0.1517	—	0.9300± 0.0716	—	15.6930± 14.2680	—	3.0831± 8.6962	—	4.6712± 9.0435	—
All data	AF-net	0.8557± 0.1072	45e-04	0.8073± 0.1430	38e-06	0.9339± 0.0678	n.s.	15.4978± 13.6157	n.s.	2.9580± 7.9512	24e-03	4.5085± 8.2928	55e-03
	Dual path	0.8599 ± 0.1074	—	0.8180 ± 0.1435	—	0.9300± 0.0701	—	15.4953± 13.6532	—	2.8910 ± 7.9537	—	4.4310 ± 8.2970	—

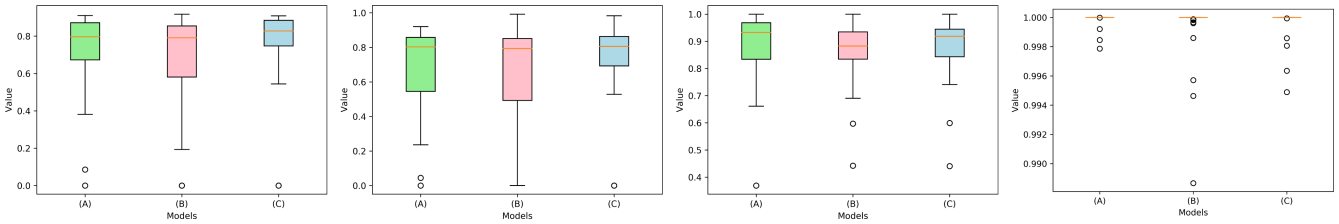
(a)

Network	DSC	Sensitivity	Precision	HD	ASSD	RMSD
U-net	0.9152 ± 0.0609	0 .9273 ± 0.0643	0 .9095 ± 0.1039	5 .3926 ± 7.7648	1 .1628 ± 0.8443	1 .7219 ± 1.6301

(b)

Network	Data w/ RVB			Data w/o RVB
	DSC	Sensitivity	Precision	Specificity
RVB-net	0.7589 ± 0.2214	0 .7339 ± 0.2290	0 .8761 ± 0.1185	0 .9999 ± 0.0005

(c)



(a) DSC (b) Sensitivity (c) Precision (d) Specificity

Module		Data w/ RVB						Data w/o RVB	
Side input	AG	DSC		Sensitivity		Precision		Specificity	
		Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value
✓		0.7162 ± 0.2260	$3.2e$-02	0.6780 ± 0.2541	$3.0e$-02	0.8763 ± 0.1322	n.s.	0.9999 ± 0.0004	n.s.
	✓	0.6831 ± 0.2527	$7.0e$-03	0.6515 ± 0.2914	$3.1e$-02	0.8617 ± 0.1177	n.s.	0.9998 ± 0.0011	n.s.
✓	✓	$\mathbf{0.7589 \pm 0.2214}$	—	$\mathbf{0.7339 \pm 0.2290}$	—	0.8761 ± 0.1185	—	0.9999 ± 0.0005	—

(e)

FIGURE 11. Ablation study of RVB-net by comparing segmentation performance of reverberation artifacts contaminated AF region between different models. (a)-(d) show the boxplots of DSC, sensitivity, precision and specificity for the three models: Baseline + Side input (in green), Baseline + AG (in pink) and RVB-net (in blue). (e) The segmentation results of the above-mentioned three models ("n.s." stands for not statistically significant).

rate policy [25]. We also applied a five-fold cross-validation method for hyperparameter tuning and validation of each neural network. In addition, all networks are initialized with a random normal initializer with a mean zero and a standard deviation of 0.001, and all the experiments are performed using a CPU (Intel CoreTM i7-6850K, 3.60GHz), 128 GB DDR4 RAM, and a four GPU (NVIDIA GTX-1080ti, 11 GB) system running Ubuntu 16.04.4 LTS.

D. EVALUATION METRICS

To evaluate the segmentation performance of different methods, several metrics are employed; DSC, sensitivity, precision, specificity, Hausdorff distance (HD), average symmetric surface distance (ASSD), and root mean squared distance (RMSD) [26], [27]. The Wilcoxon signed-rank test [28] with 5% significance level is applied to compare statistically significant differences between segmentation results.

The DSC, sensitivity, precision, and specificity are commonly-used region-based metrics to evaluate the segmentation performance, defined as:

$$\text{DSC} = \frac{2|\Omega_{GT} \cap \Omega_s|}{|\Omega_{GT}| + |\Omega_s|}, \quad \text{Sensitivity} = \frac{|\Omega_{GT} \cap \Omega_s|}{|\Omega_{GT}|}, \tag{13}$$

$$\text{Precision} = \frac{|\Omega_{GT} \cap \Omega_s|}{|O_s|}, \quad \text{Specificity} = \frac{|(\Omega_{GT} \cup \Omega_s)^c|}{|(\Omega_{GT})^c|}, \tag{14}$$

where Ω_s represents the segmented region, Ω_{GT} denotes the ground truth, and $(\cdot)^c$ is the complementary set.

Three distance-based metrics, HD, ASSD, and RMSD, are used to measure the boundary mismatches of AF

TABLE 4. Performance evaluation and comparison of AF segmentation between the proposed dual path network and other state-of-the-art semantic segmentation methods on (a) the data w/ RVB, (b) the data w/o RVB and (c) all evaluation data in (a) and (b). ("n.s." stands for not statistically significant).

	DSC		Sensitivity		Precision		HD		ASSD		RMSD	
	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value
U-net	0.8635 ± 0.0629	$1.9e$-04	0.7970 ± 0.0990	$3.2e$-05	0.9531 ± 0.0563	$3.3e$-04	13.8749 ± 7.2210	n.s.	2.6933 ± 1.4250	$3.1e$-03	4.2284 ± 2.3394	$1.2e$-02
Attention U-net	0.8750 ± 0.0717	$1.1e$-03	0.8234 ± 0.1120	$6.1e$-05	0.9466 ± 0.0637	$1.3e$-02	14.4893 ± 8.6164	n.s.	2.5732 ± 1.5321	$1.0e$-02	4.2633 ± 3.0179	$3.6e$-02
Deeplabv3+	0.8772 ± 0.0510	$1.8e$-04	0.8237 ± 0.0764	$4.8e$-05	0.9390 ± 0.0568	n.s.	11.4500 ± 6.0329	n.s.	2.0297 ± 0.7907	n.s.	3.1814 ± 1.4381	n.s.
SegNet	0.8785 ± 0.0720	$2.5e$-03	0.8569 ± 0.0873	$1.0e$-02	0.9159 ± 0.1032	$2.2e$-02	12.0288 ± 5.8331	n.s.	2.1453 ± 1.1423	n.s.	3.2699 ± 1.7428	n.s.
AF-net	0.8795 ± 0.0508	$5.0e$-04	0.8250 ± 0.0921	$1.2e$-05	0.9495 ± 0.0390	$2.0e$-02	14.9110 ± 9.9218	n.s.	2.3469 ± 1.1443	$1.4e$-03	3.7174 ± 1.9065	$4.7e$-03
Dual path	$\mathbf{0.9042 \pm 0.0361}$	—	$\mathbf{0.8863 \pm 0.0620}$	—	0.9287 ± 0.0612	—	14.4083 ± 9.9300	—	$\mathbf{1.9335 \pm 0.8870}$	—	$\mathbf{3.2230 \pm 1.6716}$	—

(a)

	DSC		Sensitivity		Precision		HD		ASSD		RMSD	
	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value
U-net	0.7450 ± 0.1690	$1.3e$-17	0.6453 ± 0.1919	$3.3e$-19	0.9578 ± 0.0800	$4.9e$-12	21.7177 ± 17.6231	$3.0e$-08	5.1537 ± 9.5700	$2.0e$-16	7.4973 ± 10.6957	$1.9e$-14
Attention U-net	0.7575 ± 0.1620	$6.9e$-19	0.6615 ± 0.1809	$8.0e$-20	0.9445 ± 0.0890	$2.9e$-07	22.5689 ± 19.0811	$9.1e$-09	5.1000 ± 9.7289	$1.1e$-16	7.5654 ± 1.1093	$2.1e$-13
Deeplabv3+	0.7787 ± 0.1326	$4.5e$-16	0.6950 ± 0.1725	$3.0e$-16	0.9311 ± 0.0925	n.s.	20.3421 ± 14.9279	$1.7e$-07	4.1501 ± 3.1770	$2.5e$-14	6.6520 ± 5.8712	$1.0e$-12
SegNet	0.7809 ± 0.1412	$2.3e$-15	0.7102 ± 0.1723	$1.8e$-14	0.9120 ± 0.1213	$1.1e$-02	19.9412 ± 15.2149	$1.3e$-06	4.1514 ± 5.4139	$1.5e$-13	6.2822 ± 7.3050	$5.5e$-10
AF-net	0.8510 ± 0.1165	n.s.	0.8030 ± 0.1517	n.s.	0.9302 ± 0.0717	n.s.	15.6175 ± 14.2120	n.s.	3.0810 ± 8.6951	n.s.	4.6652 ± 9.0421	n.s.
Dual path	$\mathbf{0.8510 \pm 0.1165}$	—	$\mathbf{0.8030 \pm 0.1517}$	—	0.9300 ± 0.0716	—	$\mathbf{15.6930 \pm 14.2680}$	—	$\mathbf{3.0831 \pm 8.6962}$	—	$\mathbf{4.6712 \pm 9.0435}$	—

(b)

	DSC		Sensitivity		Precision		HD		ASSD		RMSD	
	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value	Mean±std	p-value
U-net	0.7649 ± 0.1620	$4.7e$-22	0.6679 ± 0.1859	$6.7e$-23	0.9545 ± 0.0765	$2.6e$-13	20.4033 ± 16.5975	$5.9e$-07	4.7420 ± 8.8034	$1.3e$-18	6.9511 ± 9.8812	$2.0e$-14
Attention U-net	0.7768 ± 0.1574	$2.6e$-21	0.6892 ± 0.1877	$5.3e$-23	0.9472 ± 0.0840	$1.7e$-08	21.2210 ± 18.0054	$7.2e$-08	4.6854 ± 8.9458	$4.6e$-18	7.0140 ± 0.2820	$3.1e$-14
Deeplabv3+	0.7951 ± 0.1279	$7.0e$-19	0.7170 ± 0.1680	$3.2e$-19	0.9324 ± 0.0898	n.s.	18.8610 ± 14.1896	$6.1e$-06	3.7969 ± 3.0219	$4.4e$-15	6.0766 ± 5.5461	$2.8e$-12
SegNet	0.7970 ± 0.1368	$8.0e$-18	0.7351 ± 0.1730	$2.2e$-16	0.9128 ± 0.1176	$5.0e$-04	18.6234 ± 14.3611	$2.4e$-05	3.8200 ± 5.0170	$5.0e$-13	5.7814 ± 6.7967	$2.8e$-09
AF-net	0.8557 ± 0.1072	$4.5e$-04	0.8073 ± 0.1430	$3.8e$-06	0.9339 ± 0.0678	n.s.	15.4978 ± 13.6157	n.s.	2.9580 ± 7.9512	$2.4e$-03	4.5085 ± 8.2928	$5.5e$-03
Dual path	$\mathbf{0.8599 \pm 0.1074}$	—	$\mathbf{0.8180 \pm 0.1435}$	—	0.9300 ± 0.0701	—	$\mathbf{15.4953 \pm 13.6532}$	—	$\mathbf{2.8910 \pm 7.9537}$	—	$\mathbf{4.4310 \pm 8.2970}$	—

(c)

segmentation and the ground truth. They are all quantified in millimeters. HD is the maximum of two directed boundary Hausdorff distances in both directions, defined as:

$$\begin{aligned}&\mathrm{HD}(\Omega_{\mathrm{S}}, \Omega_{\mathrm{GT}})\\ &\quad = \max\left\{\max_{p\in\partial\Omega_{\mathrm{S}}} d(p, \partial\Omega_{\mathrm{GT}}), \max_{p'\in\partial\Omega_{\mathrm{GT}}} d(p', \partial\Omega_{\mathrm{S}})\right\},\end{aligned} \tag{15}$$

where $d(p, \partial\Omega_{\mathrm{GT}})$ is the distance between p and $\partial\Omega_{\mathrm{GT}}$. ASSD and RMSD are defined as:

$$\begin{aligned}&\mathrm{ASSD}(\Omega_{\mathrm{S}}, \Omega_{\mathrm{GT}})\\ &\quad = \frac{1}{|\partial\Omega_{\mathrm{GT}}| + |\partial\Omega_{\mathrm{S}}|}\\ &\qquad \times\Big(\sum_{p\in\partial\Omega_{\mathrm{S}}} d(p, \partial\Omega_{\mathrm{GT}}) + \sum_{p'\in\partial\Omega_{\mathrm{GT}}} d(p', \partial\Omega_{\mathrm{S}})\Big),\end{aligned} \tag{16}$$

$$\begin{aligned}&\mathrm{RMSD}(\Omega_{\mathrm{S}}, \Omega_{\mathrm{GT}})\\ &\quad = \sqrt{\frac{1}{|\partial\Omega_{\mathrm{GT}}| + |\partial\Omega_{\mathrm{S}}|}}\\ &\qquad \times\sqrt{\sum_{p\in\partial\Omega_{\mathrm{S}}} d^2(p, \partial\Omega_{\mathrm{GT}}) + \sum_{p'\in\partial\Omega_{\mathrm{GT}}} d^2(p', \partial\Omega_{\mathrm{S}})}.\end{aligned} \tag{17}$$

E. PERFORMANCE OF PROPOSED DUAL PATH NETWORK FOR AF SEGMENTATION

The AF pocket segmentation result of the dual path network, as shown in Fig. 10 and Table 3, indicated that the complementary network effectively corrected the reverberation-artifact-induced misjudgment of the AF region and provided more accurate segmentation of the AF pocket than AF-net.

As illustrated in Table 3 (a), the dual path network achieved more promising segmentation of AF pocket. Specifically, the dual path network obtained significantly better performance than AF-net on most evaluation metrics, except precision, for the image whose AF pocket is contaminated with the reverberation artifacts. Meanwhile, it yields almost no false positive errors for the image whose AF pocket is not contaminated with the reverberation artifacts.

The improvement of the dual path network is attributed to its secondary path, as shown in Table 3 (b) and (c). The accurate abdominal wall segmentation, as shown in Table 3 (b), guaranteed a satisfying generation of the derivative feature map. RVB-net achieved a promising segmentation of the reverberation artifacts, as shown in Table 3 (c). Specifically, the high precision and specificity suggested that the network did not yield fatal positive errors, while the sensitivity demonstrated that strong reverberation artifacts are correctly recognized and can compensate for AF-net, as shown in Fig. 10.

F. ABLATION STUDY OF RVB-NET

To evaluate the performance of the two modules of RVB-net, i.e., the reverberation-enriched side input and AG, by segmenting the reverberation-artifact-contaminated AF region, in this section, we performed ablation experiments on RVB-net and analyzed the effect of each module. These two modules were gradually removed from RVB-net, and the following three models were considered: (1) RVB-net, which involves three components; the baseline, which is the deeply supervised architecture, reverberation-enriched side input, and AG, (2) the baseline incorporated with

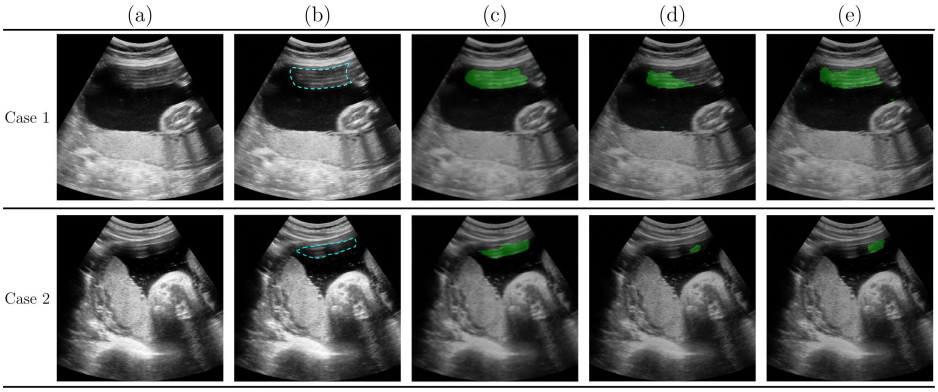


FIGURE 12. Segmentation examples of reverberation artifacts contaminated AF region for ablation study of RVB-net. (a) Input image, (b) ground truth of reverberation artifacts contamianted AF region inside the blue dotted contour. (c)-(e) show the segmentation results from RVB-net, Baseline + Side input, Baseline + AG, respectively.

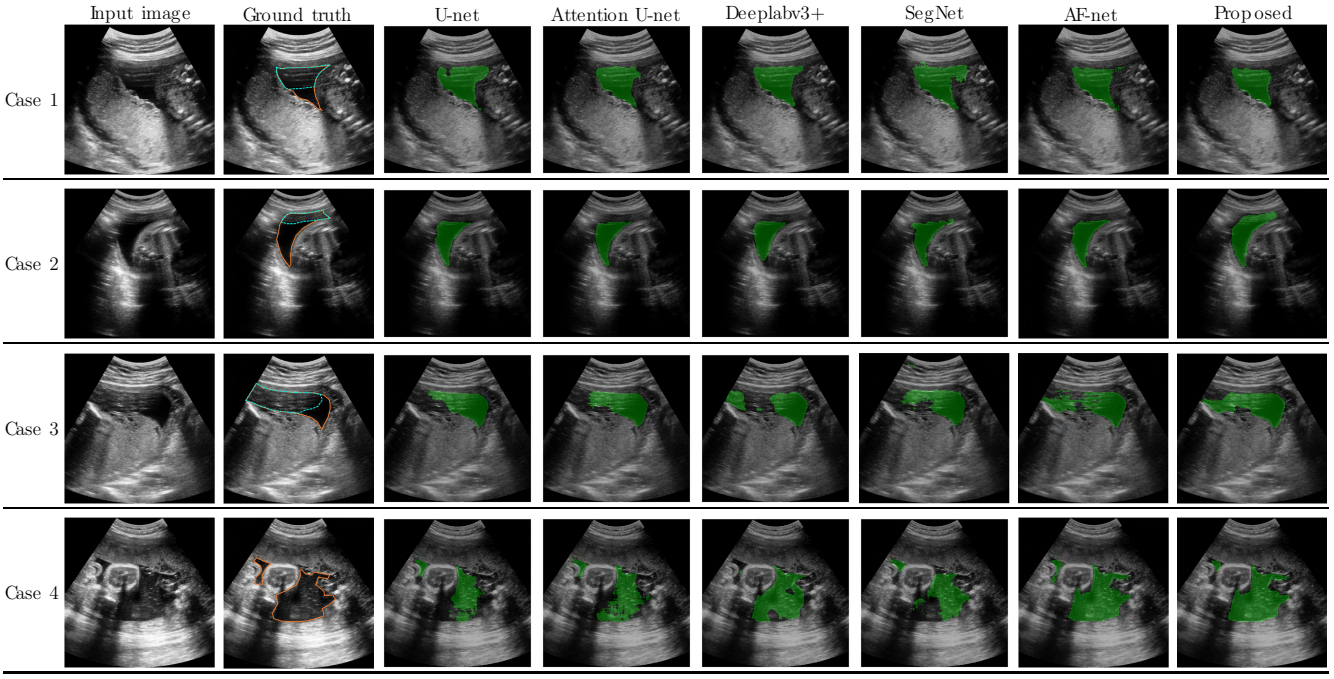


FIGURE 13. Results of AF segmentation obtained by the state-of-the-art segmentation methods and the proposed dual path network. From Case 1 to Case 3, the AF pocket is contaminated with severe reverberation artifacts. In Case 4, there is no reverberation artifacts in AF pocket. The ground truths of AF pocket and reverberation artifacts contaminated AF region are in the orange and cyan contours, respectively.

reverberation-enriched side input (denoted as "Baseline + Side input"), and (3) the baseline with AG module (denoted as "Baseline + AG"). The comparison of the segmentation performance between these models is shown in Fig. 11 and Fig. 12.

As shown in Fig. 11, integrating either the reverberation-enriched side input or AG into the baseline considerably improved the sensitivity while maintaining the precision. Hence, an improved DSC was achieved in segmenting the reverberation artifacts in the AF pocket. This indicates that our reverberation-enriched side input is an effective approach to incorporate the reverberation artifacts features into the network, and the network learned more reverberation features by using AG to guide the network to focus on the reverberation artifacts in AF pocket. When these two modules are both used, a further improved segmentation accuracy demonstrates that RVB-net takes advantage of both modules simultaneously. The segmentation results in Fig. 12 show that the Baseline + Sideinput, and the Baseline + AG models do not achieve satisfying Ω^{D}_{rvb} segmentation. Specifically, these models often fail to recognize reverberation artifacts in the AF pocket, for Case 1 to Case 2. In contrast, RVB-net provided desirable segmentation results without misjudging incorrect regions.

G. COMPARISON OF AF SEGMENTATION BETWEEN PROPOSED DUAL PATH NETWORK AND OTHER DEEP LEARNING-BASED SEGMENTATION NETWORKS

To show the advantages of proposed dual path network for accurate segmentation of AF pocket, especially for the image whose AF pockets are contaminated with reverberation artifacts, the proposed network was compared with several state-of-the-art deep learning-based segmentation methods including U-net, SegNet [29], Deeplabv3+ [25], Attention U-net [30] and AF-net.

As indicated in Table 4, the proposed dual path network outperformed other deep learning methods in AF segmentation, on the most evaluation metrics. The results in Table 4 (a) show that the dual path network had significantly better DSC, sensitivity, ASSD, and RMSD than other methods when segmenting the AF pocket contaminated with severe reverberation artifacts. Specifically, U-net achieved the highest precision but the lowest sensitivity. Although SegNet obtained a high sensitivity, the precision is not satisfying.

The qualitative comparisons between these methods are shown in Fig. 13. The first three cases (Case 1, 2, and 3) showed the advantages of our proposed method in dealing with severe reverberation artifacts in AF pocket. The proposed method captured the AF region that is contaminated with severe reverberation artifacts. On the other hand, the U-net faced difficulty in recognizing the severe reverberation artifacts in AF pocket, and SegNet misrecognized the AF region and resulted in some false positive errors. Compared to U-net, Attention U-net, Deeplabv3+, and AF-net showed slightly better performance, but they were still affected by reverberation artifacts.

The results in Table 4 (b) and the Case 4 in Figure 13 demonstrated that the dual path network and AF-net had remarkably better performance in segmenting the AF pockets not contaminated with the reverberation artifacts, compared with other methods. The evaluation of AF segmentation accuracy on all the whole evaluation data, as summarized in Table 4 (c), illustrates that the dual path network further improve the AF segmentation accuracy of AF-net and is superior to other state-of-the-art semantic segmentation networks.

IV. DISCUSSION AND CONCLUSION

This paper proposed a dual path network that was specifically designed to resolve misjudgments caused by severe reverberation artifacts when segmenting AF pockets in US images. As shown in Fig. 1, the AF area contaminated with severe reverberation artifacts can be erroneously judged as a non-AF region. Although the image characteristics (e.g., pattern, intensity, and shape) of severe reverberation artifacts are similar to those of maternal tissue layers (fat-muscle interfaces) in abdominal wall but considerably different from that of AF region, they can be recognized by appropriately utilizing the mechanism of reverberation artifacts in US imaging, which is the core idea of the proposed method.

The primary contribution of the proposed method is the development of the auxiliary network that effectively complements the AF-net by focusing on the distinct features of reverberation artifacts within the AF area. Unlike other multi-scale input layers that are generated from the input image [31], [32], the RVB-net uses the reverberation-enriched side input module to focus on features of reverberation artifacts in the local area. The AG further helps the network focus on the target area and subsequently improves the segmentation performance. This network, which was specifically designed for handling reverberation artifacts is significantly different from other deep learning-based techniques, such as [33]–[35].

Abnormal amniotic fluid amount is an important clinical cue to screen for fetal structural or genetic abnormalities or to decide on imminent delivery or intervention in cases including but not limited to intrauterine fetal growth restriction, monochorionic twin complication, or premature rupture of the membranes [36]–[38]. As the vertical distance between the anterior and posterior uterine walls represents the amniotic fluid pocket, a clear delineation of the inner contour of the anterior wall is important for the correct depth measurement [39], [40]. However, the anterior reverberation artifacts can be mistaken for anterior uterine wall or may obscure the amniotic fluid situated anteriorly, thus causing inappropriate caliper placement [40]. Such reverberation artifacts caused by highly reflective fascia, fat, and peritoneum layers in the abdominal wall abutting the enlarged uterus underneath, are frequently seen in maternal obesity and advanced gestational age during which amniotic fluid measurement is performed. The proposed method can assist the operator with uterine contour delineation in the presence of reverberation artifacts thus reducing the amniotic fluid measurement variability.

In the proposed method, the complementation process was performed by just taking union of the primary and secondary path result. Thus, the final segmentation output can be affected by errors in the secondary path. We plan to develop a new strategy of complementation that would involve the refinement between primary and secondary path results in our future work. In addition, a well-trained network by images obtained in a specific imaging setting may not work well for images obtained in other imaging settings due to different characteristic of the noise distribution. Therefore, we also plan to concern the noise disturbance in US images to improve out-of-distribution robustness for US images obtained in other imaging setting.

REFERENCES

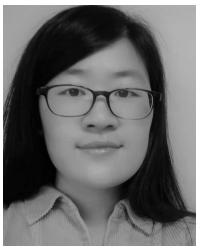
[1] F. A. Manning, L. D. Platt, and L. Sipos, "Antepartum fetal evaluation: Development of a fetal biophysical profile," *Amer. J. Obstetrics Gynecology*, vol. 136, no. 6, pp. 787–795, Mar. 1980.

[2] E. A. Dubil and E. F. Magann, "Amniotic fluid as a vital sign for fetal wellbeing," *Australas. J. Ultrasound Med.*, vol. 16, no. 2, pp. 62–70, May 2013.

[3] H. C. Cho, S. Sun, C. M. Hyun, J.-Y. Kwon, B. Kim, Y. Park, and J. K. Seo, "Automated ultrasound assessment of amniotic fluid index using deep learning," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101951.

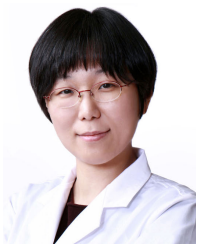
[4] J. Coombe-Patterson, "Amniotic fluid assessment: Amniotic fluid index versus maximum vertical pocket," *J. Diagnostic Med. Sonography*, vol. 33, no. 4, pp. 280–283, Jul. 2017.

[5] J. Phelan, C. V. Smith, P. Broussard, and M. Small, "Amniotic fluid vol. assessment, with the four-quadrant technique at 36-42 weeks' gestation," *J. Reproductive Med. Obstetrician Gynecolog.*, vol. 32, no. 7, pp. 540–542, 1987.
[6] J. Jang, Y. Park, B. Kim, S. M. Lee, J.-Y. Kwon, and J. K. Seo, "Automatic estimation of fetal abdominal circumference from ultrasound images," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1512–1520, Sep. 2018.
[7] L. Wu, J. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, "FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1336–1349, May 2017.
[8] B. Kim, K. C. Kim, Y. Park, J.-Y. Kwon, J. Jang, and J. K. Seo, "Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images," *Physiolog. Meas.*, vol. 39, no. 10, Oct. 2018, Art. no. 105007.
[9] H. P. Kim, S. M. Lee, J.-Y. Kwon, Y. Park, K. C. Kim, and J. K. Seo, "Automatic evaluation of fetal head biometry from ultrasound images using machine learning," *Physiolog. Meas.*, vol. 40, no. 6, Jul. 2019, Art. no. 065009.
[10] M. M. Chou, J. I. Hwang, J. J. Tseng, Y. F. Huang, and E. S. C. Ho, "Cesarean scar pregnancy: Quantitative assessment of uterine neovascularization with 3-dimensional color power Doppler imaging and successful treatment with uterine artery embolization," *Amer. J. Obstetrics Gynecology*, vol. 190, no. 3, pp. 866–868, 2004.
[11] F. M. Abu-Zidan, A. F. Hefny, and P. Corr, "Clinical ultrasound physics," *J. Emergencies, Trauma Shock*, vol. 4, no. 4, p. 501, 2011.
[12] T. Kling, K. K. Shung, and G. A. Thieme, "Reverberation reduction in ultrasonic B-mode images via dual frequency image subtraction," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 792–802, Dec. 1993.
[13] M. K. Feldman, S. Katyal, and M. S. Blackwood, "US artifacts," *Radio Graph.*, vol. 29, no. 4, pp. 1179–1189, Jul. 2009.
[14] R. De Luca, M. Bassani, L. Francalanci, F. Bertocci, F. Gelli, P. Palchetti, D. Coppini, and L. Bocchi, "A mathematical model for reverberations in biomedical ultrasound transducers: A case study," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6.
[15] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
[16] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587
[17] R. Mehta and J. Sivaswamy, "M-Net: A convolutional neural network for deep brain structure segmentation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 437–440.
[18] J. J. Dahl and N. M. Sheth, "Reverberation clutter from subcutaneous tissue layers: Simulation and *in vivo* demonstrations," *Ultrasound Med. Biol.*, vol. 40, no. 4, pp. 714–726, Apr. 2014.
[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[20] K. A. M. Said, A. B. Jambek, and N. Sulaiman, "A study of image processing using morphological opening and closing processes," *Int. J. Control Theory Appl.*, vol. 9, no. 31, pp. 15–21, 2016.
[21] P. Soille, "Opening and closing," in *Morphological Image Analysis*. Berlin, Germany: Springer-Verlag, 2004, pp. 105–137.
[22] J. Hong, B.-Y. Park, and H. Park, "Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2892–2895.
[23] S. S. Girija, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *Softw. Available From Tensorflow.Org*, vol. 39, no. 9, Mar. 2016.
[24] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467
[25] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[26] A. Fenster and B. Chiu, "Evaluation of segmentation algorithms for medical imaging," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 7186–7189.
[27] V. Yeghiazaryan and I. Voiculescu, "Family of boundary overlap metrics for the evaluation of medical image segmentation," *J. Med. Imag.*, vol. 5, no. 1, Feb. 2018, Art. no. 015006.
[28] R. Woolson, "Wilcoxon signed-rank test," in *Wiley Encyclopedia of Clinical Trials*. Hoboken, NJ, USA: Wiley, 2007, pp. 1–3.
[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[30] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.
[31] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.
[32] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3000–3009.
[33] D. A. Chanti, V. G. Duque, M. Crouzier, A. Nordez, L. Lacourpaille, and D. Mateus, "IFSS-Net: Interactive few-shot siamese network for faster muscle segmentation and propagation in volumetric ultrasound," *IEEE Trans. Med. Imag.*, early access, Feb. 9, 2021, doi: 10.1109/TMI.2021.3058303.
[34] Z. Gao, J. Chung, M. Abdelrazek, S. Leung, W. K. Hau, Z. Xian, H. Zhang, and S. Li, "Privileged modality distillation for vessel border detection in intracoronary imaging," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1524–1534, May 2020.
[35] D. Karimi and S. E. Salcudean, "Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 499–513, Feb. 2020.
[36] F. A. Manning, C. M. Hill, and L. D. Platt, "Qualitative amniotic fluid vol. determination, by ultrasound: Antepartum detection of intrauterine growth retardation," *Amer. J. Obstetrics Gynecol.*, vol. 139, no. 3, pp. 254–258, 1981.
[37] A. H. Chon, L. M. Korst, R. D. Assaf, A. Llanes, J. G. Ouzounian, and R. H. Chmait, "Midtrimester isolated oligohydramnios in monochorionic diamniotic multiple gestations," *J. Maternal-Fetal Neonatal Med.*, vol. 32, no. 4, pp. 590–596, Feb. 2019.
[38] M. Kacerovsky, I. Musilova, C. Andrys, M. Drahosova, H. Hornychova, A. Rezac, M. Kostal, and B. Jacobsson, "Oligohydramnios in women with preterm prelabor rupture of membranes and adverse pregnancy and neonatal outcomes," *PLoS ONE*, vol. 9, no. 8, Aug. 2014, Art. no. e105882.
[39] D. Gramellini, S. Fieni, C. Verrotti, G. Piantelli, D. Cavallotti, and E. Vadora, "Ultrasound evaluation of amniotic fluid volume: Methods and clinical accuracy," *Acta Bio-Medica: Atenei Parmensis*, vol. 75, pp. 40–44, Jan. 2004.
[40] S. Rutherford, C. Smith, J. Phelan, K. Kawakami, and M. Ahn, "Fourquadrant assessment of amniotic fluid volume. Interobserver and intraobserver variation," *J. Reproductive Med.*, vol. 32, no. 8, pp. 587–589, 1987.



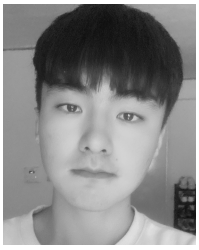
SIYU SUN is currently pursuing the Ph.D. degree with the School of Mathematics and Computing (Computational Science and Engineering), Yonsei University, South Korea. Her research interests include computer vision and deep learning in medical image analysis.

JA-YOUNG KWON received the B.S. and Ph.D. degrees in obstetrics and gynecology from Yonsei University College of Medicine, Seoul, South Korea. She was a Postdoctoral Researcher with Yale School of Medicine, New Haven, CT, USA. She is currently a Professor of obstetrics and gynecology (maternal-fetal-medicine) with Yonsei University College of Medicine. Her research interests include diagnostic ultrasound in obstetrics and AI for fetal ultrasonography.



YEJIN PARK received the master's degree in obstetrics and gynecology from Yonsei University College of Medicine, Seoul, South Korea, in 2018, where she is currently pursuing the Ph.D. degree in medicine. Her research interests include maternal fetal medicine, and deep learning for medical image analysis, especially in antenatal ultrasound.

HYUN CHEOL CHO received the B.S. degree in mathematics from Kookmin University, Seoul, South Korea, in 2018. He is currently pursuing the Ph.D. degree with the School of Mathematics and Computing (Computational Science and Engineering), Yonsei University, South Korea. His research interests include computer vision, and deep learning for medical image analysis.



CHANG MIN HYUN received the Ph.D. degree from Yonsei University, in 2021. He is currently a Postdoctoral Researcher with the School of Mathematics and Computing (Computational Science and Engineering), Yonsei University. His research interests include nonlinear and ill-posed inverse problems in medical imaging, and deep learning for medical image analysis.

JIN KEUN SEO (Member, IEEE) received the Ph.D. degree from the University of Minnesota, in 1991. He has been a Professor with Yonsei University, South Korea, since 1995. He has served as the Founding Director with the Department of the Computational Science and Engineering, Yonsei University. His research interests include inverse problems, mathematical modeling, image processing, partial differential equations, harmonic analysis, and deep learning for medical image analysis.