




Article

# Reverse Scan Conversion and Efficient Deep Learning Network Architecture for Ultrasound Imaging on a Mobile Device

Kunhyu Lee <sup>1</sup>, Min Kim <sup>2</sup>, Changhyun Lim <sup>2</sup> and Tai-Kyong Song <sup>1,\*</sup><sup>1</sup> Department of Electronic Engineering, Sogang University, Seoul 04107, Korea; gklee77@sogang.ac.kr<sup>2</sup> Hansono, Seoul 08590, Korea; drxmin@hansono.com (M.K.); limchl12@naver.com (C.L.)

\* Correspondence: tksong@sogang.ac.kr

**Abstract:** Point-of-care ultrasound (POCUS), realized by recent developments in portable ultrasound imaging systems for prompt diagnosis and treatment, has become a major tool in accidents or emergencies. Concomitantly, the number of untrained/unskilled staff not familiar with the operation of the ultrasound system for diagnosis is increasing. By providing an imaging guide to assist clinical decisions and support diagnosis, the risk brought by inexperienced users can be managed. Recently, deep learning has been employed to guide users in ultrasound scanning and diagnosis. However, in a cloud-based ultrasonic artificial intelligence system, the use of POCUS is limited due to information security, network integrity, and significant energy consumption. To address this, we propose (1) a structure that simultaneously provides ultrasound imaging and a mobile device-based ultrasound image guide using deep learning, and (2) a reverse scan conversion (RSC) method for building an ultrasound training dataset to increase the accuracy of the deep learning model. Experimental results show that the proposed structure can achieve ultrasound imaging and deep learning simultaneously at a maximum rate of 42.9 frames per second, and that the RSC method improves the image classification accuracy by more than 3%.



**Citation:** Lee, K.; Kim, M.; Lim, C.; Song, T.-K. Reverse Scan Conversion and Efficient Deep Learning Network Architecture for Ultrasound Imaging on a Mobile Device. *Sensors* **2021**, *21*, 2629. <https://doi.org/10.3390/s21082629>

Academic Editor: Arcangelo Merla

Received: 24 February 2021

Accepted: 6 April 2021

Published: 8 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** point-of-care; on-device AI; portable ultrasound; mobile system; classification

## 1. Introduction

Point-of-care ultrasound (POCUS) is an efficient tool for providing diagnostic imaging at the time and place of patient care [1]. Recently, POCUS has become more convenient to use with mobile device-based ultrasound scanners such as Healerion, Butterfly iQ (Butterfly Network Inc.), and Clarius (Clarius Mobile Health Corp.). These devices have contributed to the expansion of POCUS applications to deliver novel clinical benefits to patients [2].

With the development of this new trend, the use of ultrasound devices by unskilled and non-medical staff has become widespread [3]. As ultrasound imaging is generally performed to obtain diagnostic information in real time, POCUS users must be properly trained or technically supported at the time of ultrasound scanning. This is achievable using computer-aided diagnosis tools.

The most in-demand technical support for unskilled POCUS users might be imaging guidance, providing information on what organs are shown in each image frame [4] and whether or not the target organs are observed in the correct scan plane [5]. Automatic organ classification is essential for realizing such an imaging guidance function. Notably, object classification is also required for other advanced functions, such as automatic diagnosis or measurement of important diagnostic metrics.

Convolutional neural networks (CNNs), which made major progress in recent complex machine vision problems, have been reported to surpass human accuracy [6] in applications such as classification [7], segmentation, and object detection [8,9]. Thus, CNNs have been widely used for the classification of abdominal ultrasound images [10] and for landmark

detection of organs [11]. Thus, CNN might be the most suitable method for developing the technology for imaging guidance in POCUS.

There are two ways to implement deep learning network (DLN) inference for object classification using mobile device-based ultrasound scanners. In cloud-based inference, ultrasound images are sent to a cloud web server to compute the labels, and the results are transferred back to the mobile device. In contrast, in on-device artificial intelligence (AI), DLN inference is fully performed on a mobile device without network access [12]. Until recently, some studies have recommended the cloud computing approach using Amazon Web Services (AWS) or Azure for fast real-time implementation of CNN-based object classification, and large datasets to be easily ingested and managed to train algorithms [13,14].

However, cloud computing is not adequate for POCUS with mobile device-based ultrasound scanners, particularly when and where fast internet and wireless communication networks are not available, such as in remote areas and underdeveloped countries with poor communication infrastructure [15]. Moreover, cloud computing is vulnerable as it may attempt to violate the protection of personal information [16].

Therefore, it is advisable to implement deep learning on mobile devices without accessing the Internet. Recently, owing to rapid advances in high-performance parallel computing architecture, mobile devices are already capable of real-time software implementation of DLN using a mobile graphics processing unit (GPU). Accordingly, numerous efforts have been made to implement on-device inference using edge computing [17]. Hitherto, some efficient networks that reduce the calculation quantity and amount of memory for real-time DLN inference on a mobile device have likewise been developed [18,19].

However, in mobile device-based ultrasound scanners, entire back-end signal/image processing tasks to reconstruct images at a typical speed of 20–40 frame per second (fps) must be carried out with the same GPU. This is a major limiting factor for high-speed DLN inference in a mobile device, and to the best of our knowledge, only a few companies have developed simultaneous implementation of ultrasound back-end signal/image processing and DLN for real-time ultrasound medical imaging [20].

Notably, in ultrasound imaging, the accuracy of the DLN inference is affected by the shape of the field of view (FOV) [21]. Figure 1a shows the three most widely used types of ultrasound array transducers, namely, linear, curved linear (or convex), and sector phased arrays, where the solid arrow lines ( $S_1 \sim S_N$ ) represent the scan lines forming a single image frame [22]. Echo signals along all scan lines, which are obtained after transmission, reception, and beamforming operations along each scan direction, are stored in the echo memory in the same format as shown in Figure 1b. Image formation is achieved by organizing the lines of echo memory and processing them through a digital scan converter (DSC) that transforms them into a raster scan format for display on a video or personal computer (PC) monitor (see Figure 1c) [23].

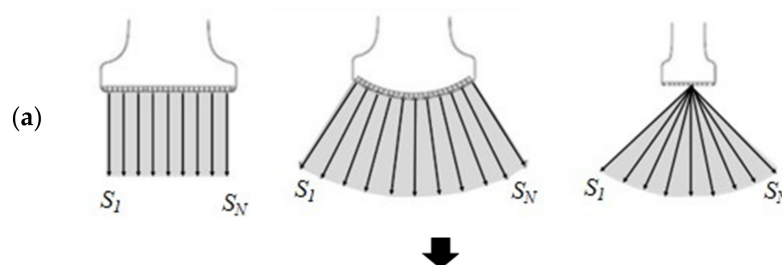
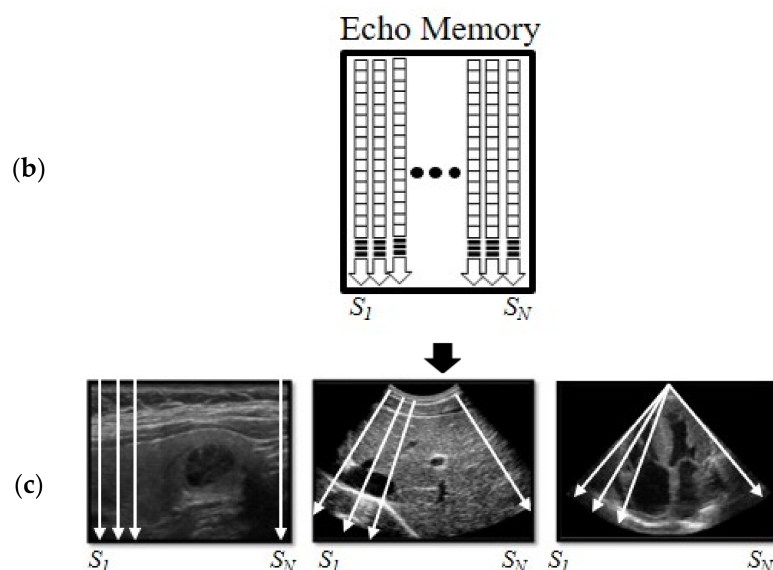


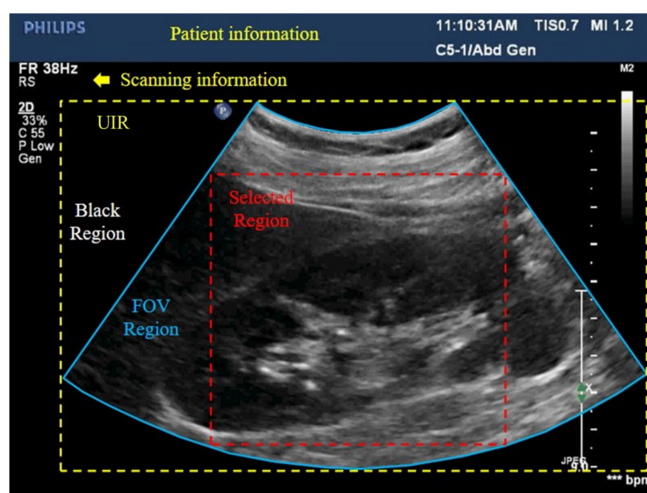
Figure 1. Cont.





**Figure 1.** (a)  $N$  scanlines made of three transducers (left panel: linear array, center panel: convex array, right panel: sector phased array.) (b) Echo memory is stored to #scanlines by #samples, regardless of type of transducers. (c)  $I_{USI}$ , digital scan-converted is reconstructed as an image with spatial information.

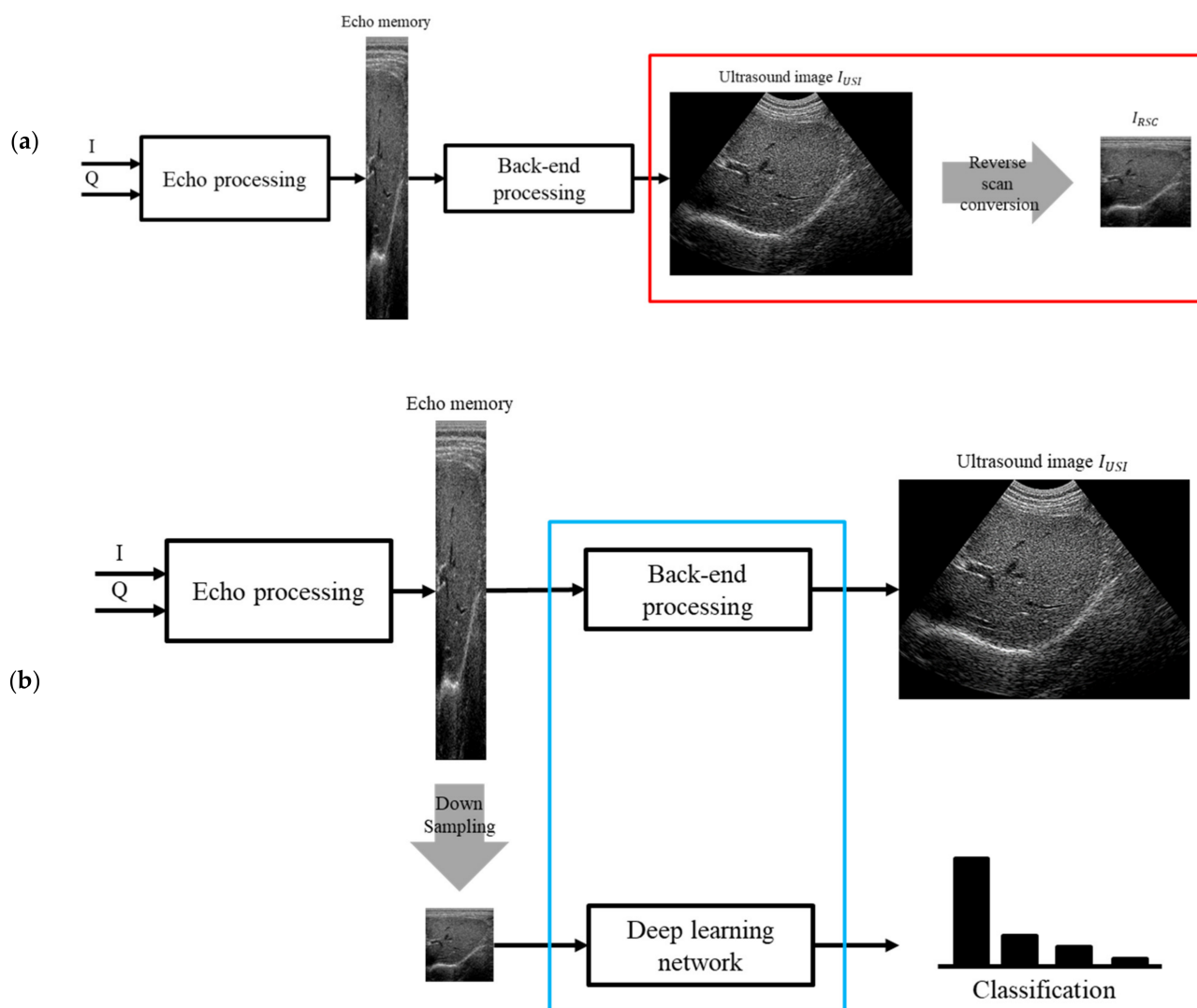
The resulting ultrasound images are displayed on the ultrasound image region (UIR), in a fixed rectangular area (yellow box in Figure 2) on a monitor screen. When the entire UIR is used as training data, not only echo information, but the FOV shape can be recognized as a feature, thus lowering the classification accuracy [21]. To avoid this, a selected region (red box in Figure 2) can be cropped within each FOV as training data. In this case, however, some objects can be partially cut off, which is another factor that lowers the classification accuracy.



**Figure 2.** Digital scan-converted ultrasound image (kidney) shown on display. This ultrasound image (USI) region includes not only the field of view (FOV) region (blue box), but also the black region according to FOV and various information. To train ultrasound images, they are generally cropped to an USI region (yellow dotted rectangle) or selected region (red dotted rectangle).

We propose an ultrasound AI edge-computing method for mobile device-based ultrasound scanners that perform ultrasound image reconstruction and DLN inference for object classification at a high frame rate required for practical ultrasound diagnosis. We propose to use a pre-DSC image, such that the entire uncropped image is used for training.

In this case, all objects in the pre-DSC image are geometrically distorted. In this approach, the DLN was trained to accurately classify the target objects in distorted images. However, no commercial ultrasound scanners provide such pre-DSC images. We refer to this procedure of obtaining pre-DSC images from DSC images as the reverse scan conversion (RSC). Figure 3a illustrates the proposed method for building pre-DSC images. We collect back-end processed  $I_{USI}$  and subsequently generate  $I_{RSC}$  using reverse-scan conversion to train the DLN.



**Figure 3.** (a) Ultrasound image formation processing structure and reverse scan conversion training data building procedure using  $I_{USI}$ , echo memory data (red box). (b) For real-time ultrasound imaging and inference, we propose a blue box structure that can effectively compute back-end processing and the deep learning network simultaneously.

In contrast, in inference (see Figure 3b), echo memory data are down-sampled and used as input to the DLN. If the DLN is implemented after back-end processing on a mobile device, it has a low ultrasonic imaging frame rate. To overcome this limitation, we propose a structure where the DLN is divided to implement ultrasound image reconstruction and DLN inference in real time. The proposed method is evaluated on a mobile device with a portable ultrasound system [3,24], and DLN models are implemented by transferring network coefficients trained in a PC workstation to a mobile device.

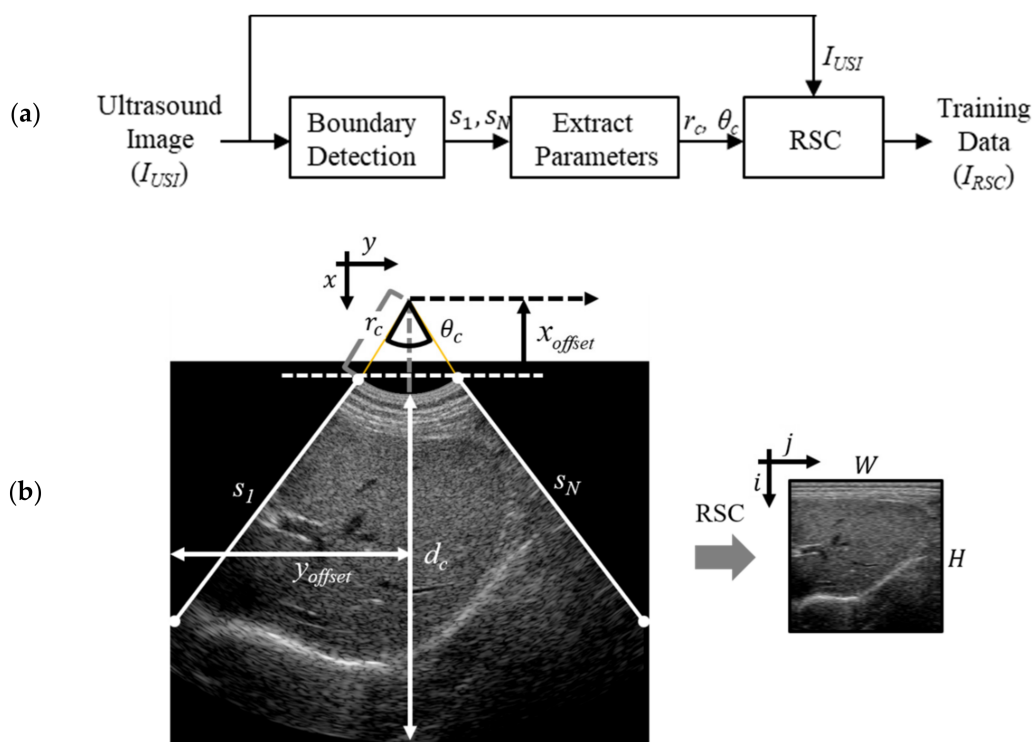
## 2. Methods

### 2.1. Reverse Scan Conversion

In diagnostic ultrasound imaging, it is customary to use a specific array transducer for a particular examination, such as a phased array for echocardiography, a convex array for the abdomen, and a linear array for the thyroid [25]. Therefore, if such conventional images are used as a training dataset, the CNN can learn the geometry of the FOV as a feature. This would negatively affect the accuracy of inference when the trained CNN is used to classify a particular organ with an input image obtained by using array transducers other than the one specific to it. This can also occur in POCUS using a mobile device-based ultrasound scanner that is usually equipped with only a one-array transducer and provides a fixed FOV shape.

Another problem with DLN inference in a mobile device-based ultrasound scanner is that high frame rate imaging may be hindered if DLN inference is followed by ultrasound back-end signal/image processing. To solve this problem, DLN inference must be possible without compromising the imaging frame rate by performing both procedures with all GPU resources. As described in the previous section, this problem can be easily solved by using the pre-DSC image data as a training dataset.

In this study, we propose an RSC method to restore the pre-DSC data from conventionally acquired and labeled images ( $I_{USI}$ ) that have been used in previous DLN studies. Figure 4 shows the processing steps used to obtain pre-DSC data.



**Figure 4.** (a) Pre-processing procedure for the training data in the ultrasound dataset.  $r_c$  and  $\theta_c$  are extracted from lines  $s_1$  and  $s_N$ . (b) Coordinate transforming a scan-converted image ( $I_{USI}$ ) to RSC image ( $I_{RSC}$ ). Left boundary, first scanline, between FOV and black region, is called  $s_1$ , whereas the right boundary, last scanline, is referred to as  $s_N$ .  $i$  and  $j$  are coordinate axes in  $I_{RSC}$ .  $I_{USI}$  has  $x$  and  $y$  coordinate axes.

In the first step, all the scanning and patient information (see Figure 2) are removed, and the UIR is converted into a binary image. Then, by detecting boundaries in the binarized UIR, the left and right FOV edges are found in the binary image to separate the FOV and black regions. Subsequently,  $N$  scanlines ( $s_1 \sim s_N$ ) are allocated over the FOV with

a uniform angle distribution:  $s_1$  represents the first scanline (i.e., left boundary) and  $s_N$  the  $N_{th}$  scanline (right boundary), as illustrated in Figure 4b for the convex array case.

In the second step of Figure 4a, the parameters to express the polar coordinates of the FOV in Figure 4b were extracted using the Hough transformation [26]. The extracted parameters include the transducer radius ( $r_c$ ), FOV angle ( $\theta_c$ ),  $x_{offset}$ , and  $y_{offset}$ .

Finally, for the pre-DSC image,  $I_{RSC}(i, j)$  is obtained in polar coordinates by applying RSC to  $I_{USI}(x, y)$  in Cartesian coordinates as follows:

$$I_{RSC}(i, j) = \mathcal{B}(\mathcal{T}(I_{USI}(x, y))), \quad (1)$$

where  $\mathcal{T}$  and  $\mathcal{B}$  represent the coordinate transformation and bilinear interpolation, respectively. Here, the coordinate transform,  $\mathcal{T}$ , is determined according to the following relationship:

$$x = (r_c + i\Delta r) \times \cos(-\theta_c/2 + j\Delta\theta) + x_{offset}, y = (r_c + i\Delta r) \times \sin(-\theta_c/2 + j\Delta\theta) + y_{offset}, \quad (2)$$

where  $\Delta r = d_c/H$  and  $\Delta\theta = \theta_c/W$ , where  $W$  and  $H$  denote the width and height of the  $I_{RSC}$ , respectively. Equation (2) for  $i$  and  $j$  is expressed as follows:

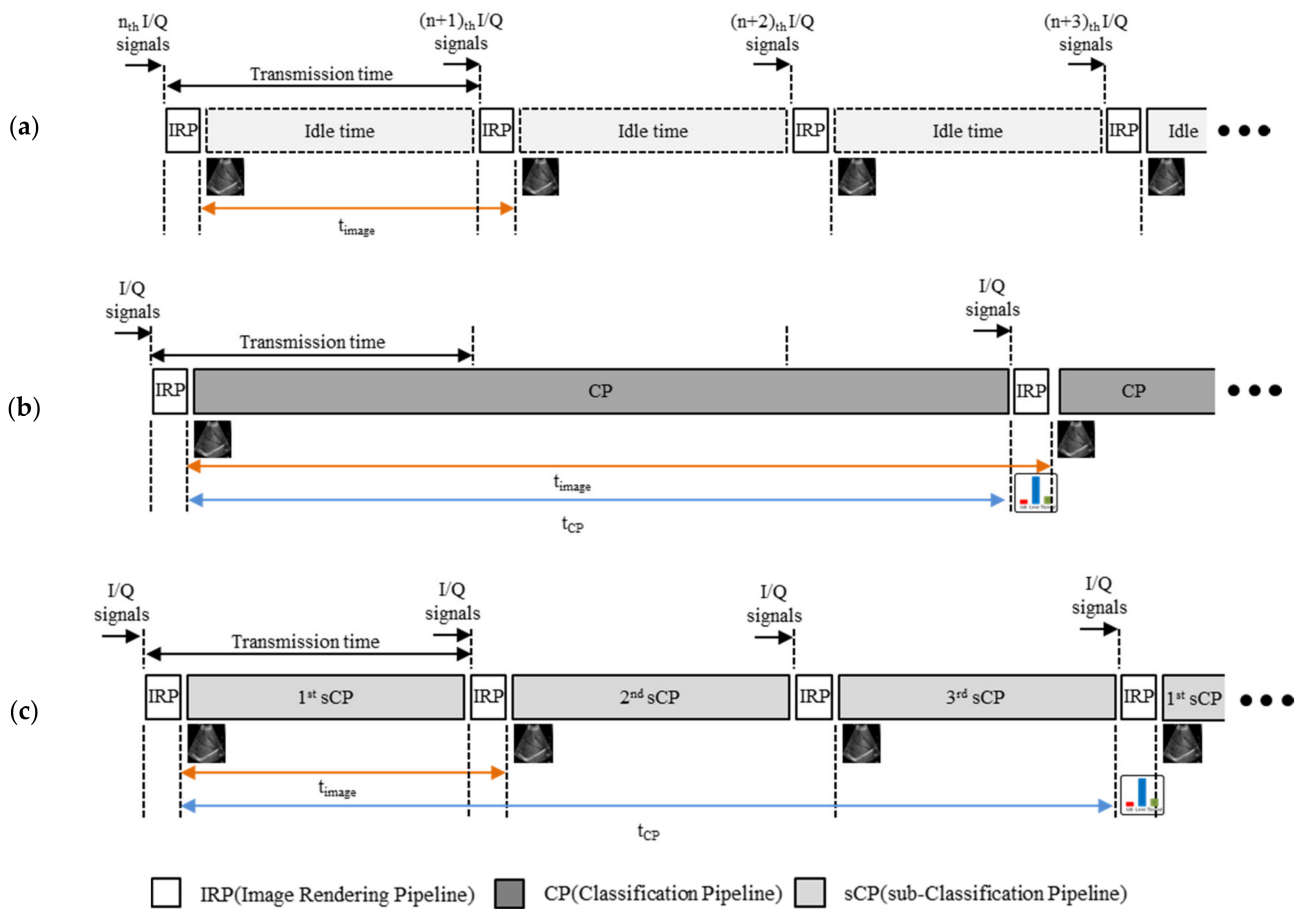
$$i = \frac{1}{\Delta\theta} \left( \arctan\left(\frac{y - y_{offset}}{x - x_{offset}}\right) + \frac{\theta_c}{2} \right), j = \frac{\sqrt{(x - x_{offset})^2 + (y - y_{offset})^2} - r_c}{\Delta r} \quad (3)$$

There are pixels (i.e., holes) that are not assigned to appropriate pixel values. Notably, the bilinear interpolation,  $\mathcal{B}$ , is employed, such that those holes must be assigned appropriate values at  $I_{RSC}(i, j)$ .

## 2.2. Structure on a Mobile Device—Frame Asynchronous Classification (FAC)

The ultrasonic signals transferred from a portable ultrasound scanner to a mobile device are in-phase/quadrature (IQ) signals. First, echo processing is applied to the IQ signals, which generates the echo memory data. Echo processing is a chain of signal processing functions, such as envelope detection and log compression. Then, the echo memory data is fed to the back-end processing block that includes the DSC. Finally, the DSC output ( $I_{USI}$ ) is displayed on the mobile device. As shown in Figure 5a, the time period to form each image frame ( $t_{image}$ ) is not determined by the processing time of the IRP ( $t_{IRP}$ ). This is because the portable ultrasound device takes time to transfer ultrasonic signals to the mobile device, and  $t_{IRP}$  is very short compared to the data transfer time. Therefore, the  $(n + 1)_{th}$  frame cannot be reconstructed immediately after the  $n_{th}$  frame is reconstructed—there is an idle time during which the IQ signals of the  $(n + 1)_{th}$  frame are being transferred.

To add a classification network in this structure, we can add a classification pipeline (CP) that implements a classification network after IRP, as shown in Figure 5b. In this structure, to implement the ultrasound image reconstruction and the classification network in real time, the entire CP must be completed within the idle time. However, the CP processing time is significantly longer than the idle time. In Figure 5b, because the entire CP is processed after the IRP, the time period to form each image frame (see orange line) is increased as much as the CP processing time (see blue line), which eventually takes longer than the data transfer time. Therefore, the ultrasound image reconstruction for the next frame is delayed. We call this the frame synchronous classification (FSC) structure [27], which has the disadvantage of delaying  $t_{image}$  as much as the classification processing time ( $t_{CP}$ ).



**Figure 5.** (a) Ultrasound image reconstruction. After image rendering pipeline (IRP) for imaging, the system must wait to receive subsequent signals due to the data transfer time. After each IRP, the ultrasound image is displayed on a mobile phone. The time period to form each ultrasound image frame is denoted by the orange bidirectional arrow. (b) frame synchronous classification (FSC) structure of a classification through deep learning with ultrasound image reconstruction in a mobile device. Classification time is denoted by the blue bidirectional arrow. (c) For real-time processing, a frame asynchronous classification (FAC) structure is made of IRP and sub-classification pipelines (sCPs).

To overcome this disadvantage of the FSC structure, we propose a structure that can implement ultrasound image reconstruction and a classification network in real time on a mobile device, as shown in Figure 5c. The key purpose of the proposed structure is to split the CP into several sub-pipelines to obtain sub-classification pipelines (sCPs) for an idle time that does not exceed the data transfer time. This structure is called frame asynchronous classification (FAC), as multiple frames are displayed during one classification result. The  $t_{CP}$  of the FAC structure is slightly longer than that of the FSC structure due to the IRPs between the sCPs. However, in real time, because the frame rate is more important than the classification rate, it does not matter whether  $t_{CP}$  is slightly longer.

Further, when the CP takes a long time to process, it can be split into a large number of sCPs. This is because  $t_{IRP}$  is very short compared to the processing time of sCPs ( $t_{sCP}$ ), as shown in Figure 5b. Therefore, even if the CP is divided into several sCPs to process ultrasound image reconstruction and DLN inference, the classification rate ( $f_{CP}$ ) has little loss. Let  $M$  denote the number of times the CP is divided into idle time, then

$$f_{image} \approx M \times f_{CP} \quad (4)$$

where  $f_{image} = 1/(t_{IRP} + t_{sCP})$ . Figure 5c shows an FAC structure in which the CP in Figure 5b is divided into three sCPs (i.e.,  $M = 3$ ).



### 3. Experiments

#### 3.1. Data

In this study, the dataset used to train the DLN comprised 38,065 frames of ultrasound images obtained from 25 volunteers, including normal and abnormal cases. Most volunteers present normal ultrasound images: there are two persons with abnormal livers, one with gallbladder problems, and one with kidney disease. Since this is not a study for diagnosing diseases, detailed abnormal cases will not be described. The abdomen is imaged, and the dataset classified into three categories: liver, kidney, and gallbladder, labeled by expert sonographers according to the guidelines specified in Reference [28]. Approximately 400 frames per organ were obtained in cine mode from various viewing angles and locations. The abdomen images of three volunteers were used as validation and test sets. A dataset comprising 3432 frames is formed by randomly dividing the cine mode of three volunteers into frames. Half of these frames were used for validation and the other half for testing. Our study and protocol were approved by the Institutional Review Board (IRB) of the Korea Centers for Disease Control and Prevention (KCDC).

#### 3.2. Embedded System

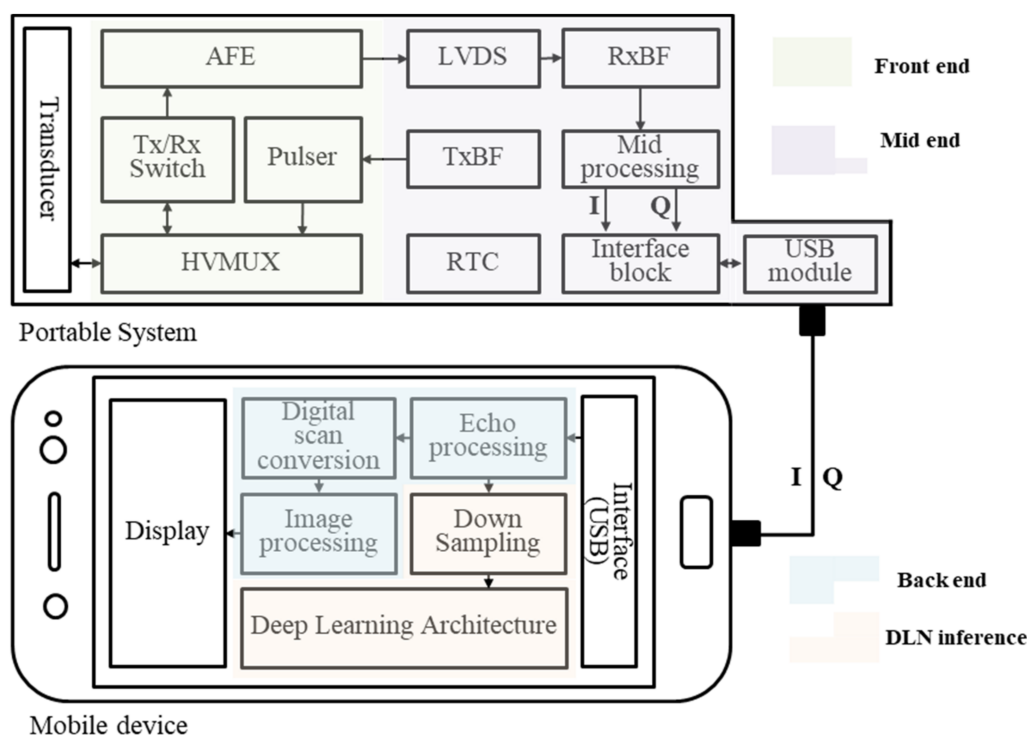
We evaluated our method and structure in a portable ultrasound imaging system-enabled smartphone (developed by our laboratory with Hansono), which consists of a smartphone (Galaxy S7, Samsung, Korea, Android 8.0) and a 32-channel system. The portable system contains analog and digital front-ends, a mid-processor, and a USB 2.0 interface, as shown in Figure 6. Further, it has a linear array and a convex array transducer attached to the system. The mobile device contains the ultrasound image reconstruction and real-time image display on a graphical user interface with the deep learning architecture. The Galaxy S7 smartphone integrates ARM Mali-T880 GPU, which has 693.6 giga floating operations per second (GFLOPS) and 4 GB memory. In the system, we used the OpenGL ES programming model [29] to harness the computing power of the mobile GPU. We used shader storage buffer objects (SSBOs) to store the DLN parameters in OpenGL. The data in SSBOs are stored in the GPU memory until the buffer is removed. In OpenGL rendering, after one IRP is finished and initialized, the following IRP must be started. However, because SSBOs are used, there is no need to upload the DLN parameters again after initialization. The calculation results are stored in the SSBOs before starting the next IRP. After rendering the next pipeline, the calculation results stored in the SSBOs can be recalled by continuing the DLN inference.

#### 3.3. Network

The deep learning model is a class of machines that learns a hierarchy of features by building high-level features from low-level ones. The CNN is a popular type of deep learning model, where trainable filters and local neighborhood pooling operations are applied in an alternating sequence, starting with raw input images. CNNs can achieve superior performance in visual object recognition and image classification tasks. CNNs have also rapidly become a methodology of choice for analyzing medical images [30], including ultrasound images. To test the method and structure proposed in this study, we explored four CNN architectures, namely, AlexNet [7], ShallowNet, MobileNet [18], and Xception [31].

AlexNet exhibits high accuracy; however, the network presented in Reference [7] is difficult to use in mobile devices due to its high computational complexity. In this work, AlexNet was used as a reference in terms of accuracy and was modified to reduce the number of nodes on the fully connected layers, from 4096 to 1024. For fast real-time computing on a mobile device, three light-weight networks (ShallowNet, MobileNet, and Xception) were chosen and evaluated. ShallowNet is a customized shallow neural network composed of two feature layers consisting of a convolution filter, max pooling, and two fully connected layers for classification. MobileNet, which is operated by a depth-wise separable convolution layer, and Xception, which uses an inception module [32], are widely

adopted light-weight networks designed to reduce computational costs while maintaining high accuracy for on-device AI.



**Figure 6.** Overall block diagram of the mobile device-based portable ultrasound system for simultaneous ultrasound signal processing and deep learning network (DLN) inference.

In the convolution layer, batch normalization and Rectified Linear Unit (ReLU) activation were applied after the convolution operation. In contrast, in the fully connected layer, we applied a 0.2 factor dropout and ReLU activation after the weighted sum. All CNNs used for evaluation are trained using stochastic gradient descent, which is commonly used for minimizing this cost function, where the cost over the entire training set is approximated with a cost of over 128 mini-batches of data. A learning rate of  $10^{-6}$  ensured proper convergence for all four networks. A smaller learning rate slowed down the convergence, and a larger learning rate often caused convergence failures. The RSC and USI datasets were trained and evaluated for each network. The datasets are stored in  $800 \times 600$  DSC format, and pre-processing must be performed to reduce their size to  $64 \times 64$ , which is used as input to the CNN. To verify the accuracy of the RSC method, the dataset of the USI region was trained separately and compared. We confirmed the performance of the proposed structure through DLN inference on a mobile device.

The parameters and calculations of the CNNs used in this study are presented in Table 1. The training process was performed in Keras with the TensorFlow framework [33] using an NVIDIA GeForce GTX 1080Ti (11 GB on-board memory) on Windows 10 for 200 optimization epochs with unit Gaussian random parameter initializations [34].

**Table 1.** Number of parameters and calculations in each network.

	Parameters	Calculations
AlexNet	7.1 M	305.3 M
ShallowNet	2.1 M	22.2 M
MobileNet	3.2 M	44.3 M
Xception	16.0 M	25.7 M

#### 4. Results

We evaluate the effectiveness of building an ultrasound image training dataset using the RSC method, as well as the performance of the FAC structure that executes ultrasound image reconstruction in real-time with DLN inference. Because the FAC structure can divide the CP to operate without delay between IRPs, reconstruction processing with classification is possible on the mobile device in real time.

When calculating the network, the system calculates two-dimensional (2D) convolution in the convolution layers and a weighted sum in the fully connected layers. For minimal memory input/output (I/O) access, the area for the convolution calculation is stored in the local (shared) memory. The convolution operation is calculated in the form of a weighted sum, and the calculation of fully connected layers is processed through the reduction process using local memory.

##### 4.1. RSC without FOV Dependency

The performance of the four networks was evaluated according to how ultrasonic images were built into a training dataset. Each of the four different networks were trained on the USI and RSC datasets and trained to predict three different organs. Table 2 shows the accuracy of each network that classifies abdominal ultrasound images, including the liver, gallbladder, and kidney. AC, SE and SP of USI and those of RSC were compared and the higher values were marked in bold. To quantitatively evaluate the RSC method, the accuracy, sensitivity, and specificity were used as performance measurements, which are defined as follows.

$$\begin{aligned} accuracy &= (TP + TN) / (TP + FN + TN + FP), \\ sensitivity &= TP / (TP + FN), \\ specificity &= TN / (TN + FP), \end{aligned} \quad (5)$$

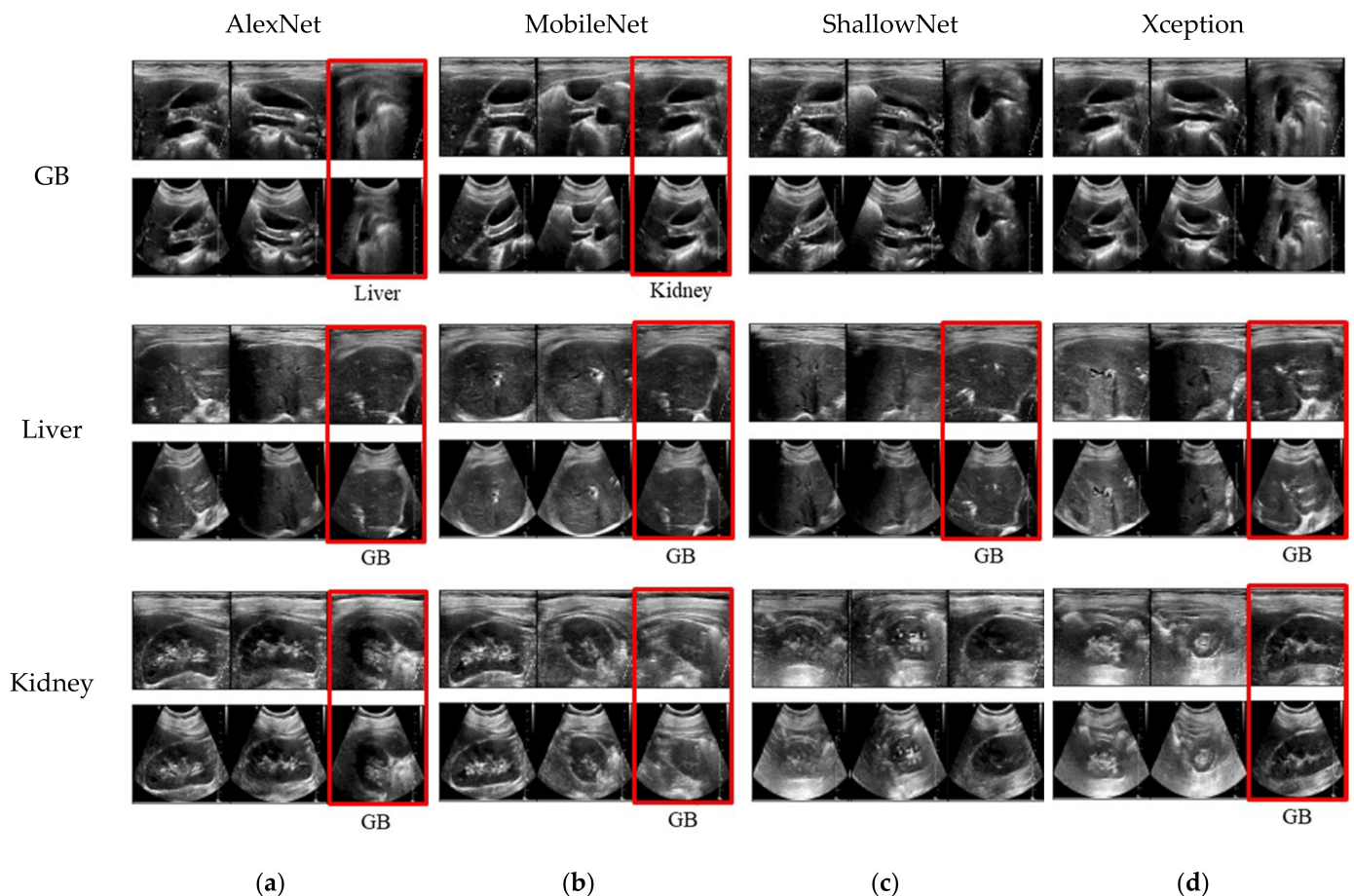
where  $TP$ ,  $FN$ ,  $TN$ , and  $FP$  represent the true positive, false negative, true negative, and false positive, respectively. For example, let the ground truth be liver, if the model prediction is liver, then it is judged as positive; otherwise, it is judged as negative. The results of networks trained with the RSC training dataset used in the experiment were more accurate than those in the network trained using USI data. The USI training dataset contains less information than the RSC training dataset at the same resolution. Furthermore, the networks trained by the USI training dataset can learn the FOV of ultrasound images as a feature, and thus they are less accurate.

Figure 7 shows the results classified for each trained network. The rows corresponding to each class are presented as pairs of RSC and USI datasets made from the same DSC images.

**Table 2.** Comparison of classification accuracies according to methods of building training data.

	USI			RSC		
	AC	SE	SP	AC	SE	SP
AlexNet	93.90	94.29	96.99	<b>98.55</b>	<b>95.94</b>	<b>97.90</b>
MobileNet	62.72	51.99	76.54	<b>68.73</b>	<b>55.56</b>	<b>78.72</b>
ShallowNet	84.00	50.65	76.42	<b>94.17</b>	<b>73.62</b>	<b>87.47</b>
Xception	88.96	67.42	<b>84.34</b>	<b>93.69</b>	<b>74.33</b>	84.32
Average	85.53	66.09	83.57	<b>88.79</b>	<b>74.86</b>	<b>87.10</b>

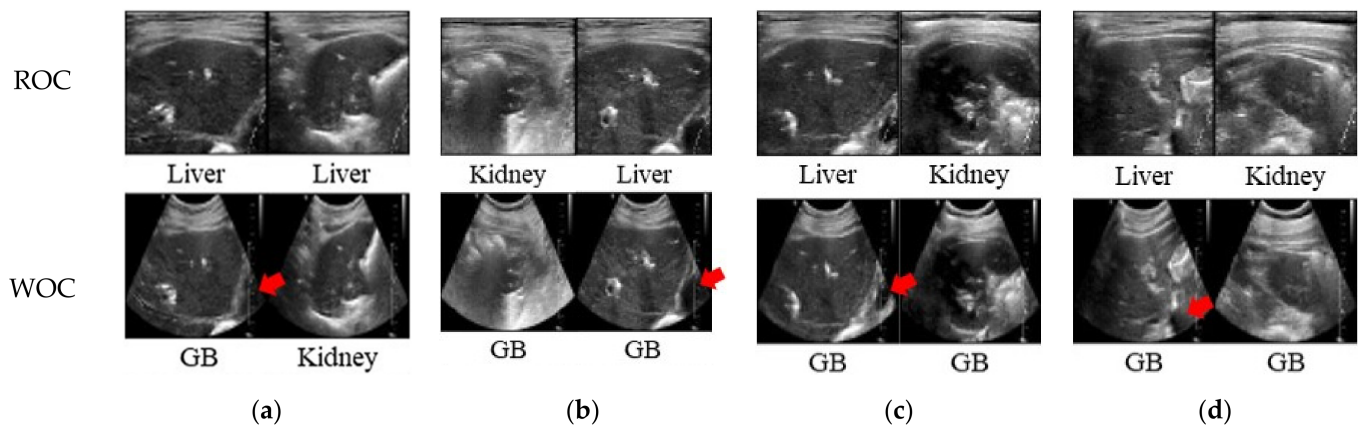
AC, accuracy; SE, sensitivity; SP, specificity.



**Figure 7.** Image pairs categorized from abdomen to gallbladder, liver, and kidney. The upper row in each organ is the result obtained by the network trained by the RSC method, and the lower row is obtained by the network trained by the USI training dataset with down-sampled DSC images. Each column depicts the result of (a) AlexNet, (b) MobileNet, (c) ShallowNet, and (d) Xception. Red boxes denote errors in the results of USI training dataset, and a misclassified organ is specified below the image.

Figure 8 shows samples of images that are trained in each network by the RSC and USI datasets, which are classified correctly and incorrectly. Of these, Figure 7 shows a pair of images that were correctly classified when training with an RSC dataset but were misclassified when training with a USI dataset.

When a feature is not clearly revealed in the image, the network trained with the USI dataset is misclassified, whereas the network trained with the RSC dataset can be classified accurately. The most common misclassified characteristic is the recognition of the liver or kidney as the gallbladder when trained with the USI dataset. The gallbladder is a pear-shaped sac, resting on the underside of the right portion of the liver [35]. In addition, the gallbladder appears anechoic, showing nothing inside the thin walls in the case of a healthy person. If a pear-shaped and anechoic structure is observed in the liver or kidney, the network trained from the USI dataset may mistake the liver or kidney as the gallbladder. This is because the amount of information varies with axial depth. Generally, there is less information at the near than at the far depth. In Figure 8, the anechoic part of the misclassified image is located under the liver. The fact that the RSC dataset has a larger FOV region than the USI dataset also affects how accurately the organs can be classified in the abdomen.



**Figure 8.** Image pairs correctly classified in the network trained by the RSC method, and misclassified in the network trained by the without cropping (WOC) method. Red arrows indicate anechoic region at deep depth, which causes misclassification. (a) AlexNet, (b) MobileNet, (c) ShallowNet, and (d) Xception.

#### 4.2. Structure for Real-Time Processing

To verify the proposed structure on the smartphone, after training various CNNs with the RSC dataset in a PC workstation, trained coefficients were transferred to a smartphone and used for DLN inference.

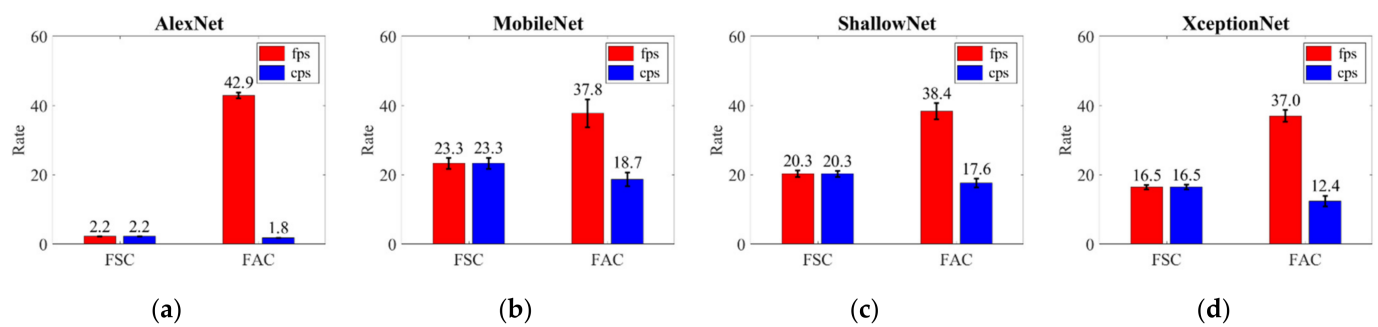
Table 3 shows how long it takes to operate DLN inference in each network when processed in the FSC structure in Figure 5b and the number of sCPs required for real-time processing in the FAC structure in Figure 5c. AlexNet has the longest processing time and the most sub-pipelines, because it has the largest number of calculations used in this study. Compared to Xception, MobileNet has a higher number of calculations, but the processing time is shorter, because the number of layers in Xception is significantly higher than that of MobileNet.

**Table 3.** Processing time with ultrasound back-end process in each network and the number of dividing for real-time processing.

		Processing Time (ms)	#Dividing
CP	AlexNet	447	22
	MobileNet	43	2
	ShallowNet	49	2
	Xception	61	3
IRP		4	(not divided)

Figure 9 shows the results of frame and classification rates by DLN inference for classification in a mobile device with ultrasound image reconstruction. With little reduction, we can increase the frame rate for real-time ultrasound image reconstruction with DLN inference in the FAC structure rather than FSC. The larger the quantity of calculations in the CNN, the more sub-pipelines we must divide, such that the growth rate of the frame rate also increases. It is effective to increase the frame rate even when the classification rate decreases. This is because ultrasound image reconstruction must be processed in real time, rather than DLN inference.





**Figure 9.** Result of ultrasound image frame rate (fps) and classification rate (cps, classification per second) during ultrasound image reconstruction with DLN inference. Frame and classification rates are represented as bar graphs with average value and variance according to frame synchronization in (a) AlexNet, (b) MobileNet, (c) ShallowNet, and (d) XceptionNet.

In Figure 9, the variance of the result of the FAC structure is usually larger than that of the FSC structure. This is because we did not divide all networks equally. When there are  $N$  sCPs, the sum of the calculation times at each sub-pipeline is not exactly the same as the entire CP operating time. After the first sCP operation of the  $N$  sCP, the network is terminated at an appropriate point in the spare time. Even if a time limit offered by users is given, it is impossible to quit the running task and run other tasks when calling the OpenGL Shader. Although the following rendering must be called during the operation, it may not be immediately run, and a delay may occur.

To reduce the variance of the operation time, the most obvious method is to analyze the performance time for each shader and divide the network accurately. It is necessary to run only the task that ends before the subsequent rendering operation. However, if the runtime is strictly limited to ensure the following rendering task, the network will have to be divided into smaller pipelines, and the idle interval of the GPU will increase. This process is not effective for executing ultrasound image reconstruction using DLN inference. Each Shader operation was also designed to maximize the occupancy, such as loading the adjacent area of the image to the GPU shared memory for memory access order and efficiency of operation. Therefore, the variance of the result is inevitable, because it is intended to ensure the efficiency of the DLN inference and ultrasonic image reconstruction.

## 5. Discussions and Conclusions

We proposed an RSC method to build a training dataset for accurate training of the CNN with ultrasound images and a structure to perform ultrasound image reconstruction with DLN inference in real time on a mobile device. To evaluate the proposed RSC method, we compared the accuracy of the CNNs trained with the RSC and USI training datasets. The average accuracy of the CNNs trained with the training dataset generated by the RSC method was approximately 3% higher than that of the training dataset of the USI region. Furthermore, the proposed structure was evaluated on a portable ultrasound device and smartphone, and the frame rate was improved by 77%.

However, we think the proposed method has some limitations. First, the proposed method was validated using a mobile device-based ultrasound scanner that is equipped with only a one-array transducer and provides a fixed FOV shape. RSC is not required for devices that support multi-FOVs with one array, such as Butterfly IQ. Second, although both the RSC process for training and the down-sampling process for DLN inference employ bilinear interpolation, the pre-DSC image obtained by down-sampling the IQ signal data for the DLN inference and the  $I_{RSC}$  used in the training process are different. This is because the geometry of the  $I_{RSC}$  is changed during the DSC and RSC processes. Therefore, further work is necessary to investigate the effect of such differences on the inference accuracy and how to improve the performance. Finally, we designed networks that classify the three major organs of the abdomen, but we have not included a category for ‘unsure’ or ‘nothing’ images, in which three organs are not included nor accurately

classified. For practical applications, the proposed method should be improved to deal with these two categories.

The RSC method of building a training dataset may improve the accuracy of various applications in training using ultrasound images. Further, the real-time structure has the advantage that ultrasound image reconstruction with DLN inference can be made without reducing the frame rate of ultrasound imaging. Using the proposed method and structure in this study, a guide for unskilled people not familiar with ultrasound imaging is effectively provided through improved POCUS.

**Author Contributions:** Conceptualization T.-K.S. and M.K.; methodology, K.L., M.K., C.L. and T.-K.S.; software, K.L. and C.L.; formal analysis, K.L., M.K. and T.-K.S.; investigation, K.L.; writing—original draft preparation, K.L. and T.-K.S.; writing—review and editing, K.L., M.K. and T.-K.S.; visualization, K.L.; supervision, K.L. and T.-K.S.; project administration, T.-K.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (9991007022).

**Institutional Review Board Statement:** This study was approved by the Institutional Review Board of the Korea Centers for Disease Control and Prevention (KCDC) (4-2017-1117).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Moore, C.L.; Copel, J.A. Point-of-care ultrasonography. *N. Eng. J. Med.* **2011**, *364*, 749–757. [[CrossRef](#)] [[PubMed](#)]
- Ailon, J.; Mourad, O.; Nadjafi, M.; Cavalcanti, R. Point-of-care ultrasound as a competency for general internists: A survey of internal medicine training programs in Canada. *Can. Med. Educ. J.* **2016**, *7*, e51. [[CrossRef](#)] [[PubMed](#)]
- Levin, D.C.; Rao, V.M.; Parker, L.; Frangos, A.J. Noncardiac point-of-care ultrasound by nonradiologist physicians: How widespread is it? *J. Am. Coll. Radiol.* **2011**, *8*, 772–775. [[CrossRef](#)] [[PubMed](#)]
- Huang, Q.; Zhang, F.; Li, X. Machine learning in ultrasound computer-aided diagnostic systems: A survey. *BioMed Res. Int.* **2018**, *2018*, 5137904. [[CrossRef](#)] [[PubMed](#)]
- Baumgartner, C.F.; Kamnitsas, K.; Matthew, J.; Fletcher, T.P.; Smith, S.; Koch, L.M.; Kainz, B.; Rueckert, D. SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans. Med. Imaging* **2017**, *36*, 2204–2215. [[CrossRef](#)] [[PubMed](#)]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
- Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
- Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
- Cheng, P.M.; Malhi, H.S. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J. Digit. Imaging* **2017**, *30*, 234–243. [[CrossRef](#)] [[PubMed](#)]
- Xu, Z.; Huo, Y.; Park, J.; Landman, B.; Milkowski, A.; Grbic, S.; Zhou, S. Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 711–719.
- Guo, T. Cloud-based or on-device: An empirical study of mobile deep inference. In Proceedings of the 2018 IEEE International Conference on Cloud Engineering (IC2E), San Francisco, CA, USA, 2–7 July 2018; pp. 184–190.
- Song, I.; Kim, H.-J.; Jeon, P.B. Deep learning for real-time robust facial expression recognition on a smartphone. In Proceedings of the 2014 IEEE International Conference on Consumer Electronics (ICCE), Taipei, Taiwan, 26–28 May 2014; pp. 564–567.
- Hosseini, M.-P.; Soltanian-Zadeh, H.; Elisevich, K.; Pompili, D. Cloud-based deep learning of big EEG data for epileptic seizure prediction. In Proceedings of the 2016 IEEE Global Conference on Signal and Information Processing, Greater Washington, DC, USA, 7–9 December 2016; pp. 1151–1155.

15. Center, P.R. Internet seen as positive influence on education but negative on morality in emerging and developing nations. *Pew Res. Cent.* **2015**, 13–20.
16. Hon, W.K.; Millard, C.; Walden, I. The problem of ‘personal data’ in cloud computing: What information is regulated?—The cloud of unknowing. *Int. Data Priv. Law* **2011**, *1*, 211–228. [[CrossRef](#)]
17. Liu, L.; Li, H.; Gruteser, M. Edge assisted real-time object detection for mobile augmented reality. In Proceedings of the 25th Annual International Conference on Mobile Computing and Networking, Los Cabos, Mexico, 21–25 October 2019; pp. 1–16.
18. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
19. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
20. Liu, J.Y.; Xu, J.; Forsberg, F.; Liu, J.-B. CMUT/CMOS-based butterfly IQ-A portable personal sonoscope. *Adv. Ultrasound Diagn. Ther.* **2019**, *3*, 115–118.
21. Madani, A.; Ong, J.R.; Tibrewal, A.; Mofrad, M.R. Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit. Med.* **2018**, *1*, 1–11. [[CrossRef](#)] [[PubMed](#)]
22. Szabo, T.L. *Diagnostic Ultrasound Imaging: Inside Out*; Academic Press: Cambridge, MA, USA, 2004.
23. McGahan, J.P.; Goldberg, B.B. *Diagnostic Ultrasound: A Logical Approach*; LWW: Philadelphia, PA, USA, 1998.
24. Kim, J.H.; Yeo, S.; Kim, M.; Kye, S.; Lee, Y.; Song, T. A smart-phone based portable ultrasound imaging system for point-of-care applications. In Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 14–16 October 2017; pp. 1–5.
25. Angelsen, B.A.; Torp, H.; Holm, S.; Kristoffersen, K.; Whittingham, T.A. Which transducer array is best? *Eur. J. Ultrasound* **1995**, *2*, 151–164. [[CrossRef](#)]
26. Duda, R.O.; Hart, P.E. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM* **1972**, *15*, 11–15. [[CrossRef](#)]
27. Burgos-Artizzu, X.P.; Coronado-Gutiérrez, D.; Valenzuela-Alcaraz, B.; Bonet-Carne, E.; Eixarch, E.; Crispi, F.; Gratacós, E. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Sci. Rep.* **2020**, *10*, 10200. [[CrossRef](#)] [[PubMed](#)]
28. Palmer, P.E. *Manual of Diagnostic Ultrasound*; World Health Organization: Geneva, Switzerland, 1995.
29. Shreiner, D.; Sellers, G.; Kessenich, J.; Licea-Kane, B. *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Version 4.3*; Addison-Wesley: Boston, MA, USA, 2013.
30. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciampi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
31. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
32. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
33. Chollet, F. Keras. Available online: <https://github.com/keras-team/keras> (accessed on 8 April 2021).
34. Wan, L.; Zeiler, M.; Zhang, S.; Le Cun, Y.; Fergus, R. Regularization of neural networks using dropconnect. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1058–1066.
35. Stocksley, M. *Abdominal Ultrasound*; Cambridge University Press: Cambridge, UK, 2001.