International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 32, Issue 01) and (Publishing Month: July 2016)
(An Indexed, Referred and Impact Factor Journal)
ISSN: 2319-6564
www.ijesonline.com

ECHOCARDIOGRAPHIC IMAGE SEGMENTATION USING DEEP RES-U NETWORK

Kisan Prayag Patro[1]
Aryan Institute of Engineering & Technology, Bhubaneswar, Odisha

Prakash Kumar Sarangi[2]
NM Institute of Engineering and Technology, Bhubaneswar, Odisha

Bhaktipadarbinda Mohanty[3]
Raajdhani Engineering College, Bhubaneswar, Odisha

ABSTRACT

Cardiac function assessment using echocardiography is a crucial step in daily cardiology. However, cardiac boundary segmentation and in particular, ventricle segmentation is a challenging procedure due to shadows and speckle noise. Manual segmentation of the cardiac boundary is a time-consuming process which rules out con- ventional segmentation for many situations such as emergency cases and image-guided robotic interventions. Therefore, providing an efficient and robust autonomous segmentation method is crucial for such applications. In this paper, a fast and fully automatic deep learning framework for left ventricle segmentation is proposed. This model couples the advantages of ResNet and U-Net to provide reliable segmentation results. We propose a new encoder in the U-Net, defined as ResU which is a modified version of ResNet-50 and has a superiority over ResNet in data denoising. We trained this model on the dataset CAMUS (Cardiac Acquisitions for Multi-structure Ul- trasound Segmentation) which is a large, publicly available and fully annotated dataset for 2D echocardiographic assessment. It is shown that this model outperforms other state-of-the-art methods in terms of accuracy with a Dice metric of 0.97 ± 0.01.

1. Introduction

Cardiac imaging is an important tool in evaluating cardiac functionality and image-guided cardiac interventions. Echocardiography is the most cost-effective image modality when compared with MRI and CT [1]. Also, echocardiography devices are available in portable versions that can examine patients' heart functionality outside of clinics [2]. Moreover, the ability to provide real-time images of the heart is another essential advantage of echocardiography scanners. However, echocardiography images have several issues that affect the results of segmenting the cardiac boundaries, and in particular those of the left ventricle. These images, in addition to shadows and dropouts, have multiplicative noise speckles which happen as a result of reflecting the sound wave echoes to the transducer.

Among many applications of cardiac imaging, image-based cardiac evaluations in emergency cases and interventions require real-time segmentation of the endocardial borders. Nevertheless, conventionally-used manual segmentation of the cardiac boundary is a time-consuming process and prone to poor reproducibility. Hence, automated segmentation methods are required in this field to provide faster cardiac functional analysis. Traditional segmentation methods, such as thresholding and region-based segmentation, are not reliable solutions for finding cardiac boundaries and require a pre-processing stage such as removing multiplicative noise [3]. Several analytical methods have also been proposed in the literature [4–7]. However, these methods, in general, are either computationally expensive and semi-autonomous or require stringent constraints to provide a correct estimation.

Accordingly, an automatic and real-time segmentation method of the endocardium borders is beneficial for interventional procedures and intensive care unit applications with monitoring requirements.

Numerous echocardiographic segmentation techniques have appeared in the literature, each of which tackles the accompanying ultrasound issues differently. The final contour produced depends on several factors (such as the initialization position, the used detection method, and data quality), and estimating the endocardial border accurately and consistently remains a challenging task. Conventional intensity gradient-based methods are not recommended in this field because they have limited success for clinical images [8]. On the other hand, statistical models that learn offline shapes have received considerable attention in echocardiographic image segmentation, especially after the work presented by Cootes et al. [9]. These models provide a significant advantage in the form of motion priors, giving the tracking process robustness against some echocardiographic image issues such as

shadows. But such models have a significant limitation, particularly pertaining to the assumption that different patients have similar cardiac structures. This assumption may not hold for new images due to subject tissue variations. This could be compensated for by providing a better initialization shape. Recently, deep learning techniques have garnered much attention in the field of computer vision due to their speed and promising accuracy. These methods are used in several applications, such as object detection, pose estimation and, most importantly, in object segmentation [10,6,5]. However, deep learning segmentation methods require an extensive and representative amount of annotated data to provide reliable segmentation results. This process requires an expert to delineate the endocardium border in potentially thousands of images, a process that is highly tedious and time-consuming. Also it should be noted that among cardiac chambers, left ventricle (LV) is responsible for pumping blood to the systemic circulation. As a result, most available cardiac segmentation methods were designed to track this chamber [11].

One of the obstacles on using deep learning methods for the ventricles segmentation was the absence of a sufficient and reliable dataset that could be used for training. However, Leclerc et al. [10] provided a labeled dataset which offers a variety of echocardiograms with different qualities (good, medium, and poor). Furthermore, to make a reliable dataset, Leclerc et al. [10] included echocardiograms that were contaminated with shadows and dropouts. However, in their paper, poor quality echocardiograms were ignored during the training process.

In the recent literature, some approaches have been proposed to use a deep neural network for segmenting the endocardium border of LV or identifying the viewpoint. For instance, Carneiro et al. [7,12] developed an automated method that uses deep learning to track the cardiac boundary. Gao et al. [13] reported another deep learning technique which classifies the echocardiography viewpoint. Leclerc et al. [10] compared the state-of-the-art of non-deep-learning methods, and encoder–decoder-based architectures showing the superiority of deep learning methods to their counterparts. Their provided CAMUS dataset comprises of four- and two-chambers acquired from 500 patients with manual segmentation (references) of the LV_{Endo}, myocardium and left atrium (LA).

A deep learning method is introduced in this paper to segment cardiac boundaries efficiently without manual initialization. This network has been trained using echocardiographic images with their segmentations done manually to segment new images that were not present in the training dataset. This network is designed based on U-Net [14] and a modified residual network (ResNet) [15] to enable having significant number of layers and enhanced accuracy. U-Net is known for its applicability of producing a higher accuracy in image segmentation applications, especially, in medical image segmentation studies. On the other hand, ResNet has several advantages such as accelerating the training speed of the networks, and reducing the effect of vanishing gradient. As a result, ResNet increases the network depth to reach over 100 layers. Also, ResNet obtains higher accuracy in image classification.

In this paper, we propose a new hybrid network for deep learning to improve information preservation. Accordingly, we also propose different training process. In this context, the purpose of this paper is to provide answers to the following three questions:

(1) How much improvement has our model contributed to the segmentation of echocardiograms?
(2) How efficient is our design and is it applicable to real-time applications?
(3) Will multi-stage gradual training process help decrease the training time?

2. Related work

Similar to any image segmentation technique, the results of echocardiographic image segmentation depend on the data quality. With the presence of artifacts (shadows, noise, and dropout), the segmentation process becomes a complicated task. There are numerous segmentation approaches and just a few are considered when segmenting echocardiographic images. Among those methods, active shape model (ASM) technique [9] provides notable results in mitigating some of the aforementioned artifacts. It is also claimed that ASM can segment echocardiographic frames better than its counterparts that place their decision making on intensity values. In general, segmentation methods treat the procedure as calculating the probability of having a correct segmentation using given information represented in training data (images, labels).

Most of the available segmentation methods consist of two critical stages [16]. The first is the initialization step, which provides the segmentation method with a starting shape. The second is the process of searching for the optimum boundary, using the neighbourhood of the initial shape. Methods such as ASM and active contours require a manual initialization to start the segmentation process. Also, the accuracy of the final segmentation results depends on the initialization stage. The initialization process in ASM requires placing an average shape (acquired by training) on the target image to start the segmentation process. However, moving this initialization shape several pixels in any direction will produce a different result. Therefore, in this case, a method which produces the same results is crucial to avoid confusion and to remove manual initialization.

From this perspective, and to have real-time results, deep learning methods have gained popularity. Carneiro et al. [7] developed a deep learning method to segment LV_{Endo} into 2D echocardiographic images. However, their method tends to misdetect the middle part of the left wall of the LV. Their objective was to find the LV contour using the following decision function:

$$s = E_s[I] = \int_s sp_s(\tilde{I}) \, ds \quad (1)$$

The outline of this paper is as follows. An overview of cardiac boundary segmentation methods is detailed in Section 2. The proposed network design is outlined in Section 3, followed by simulations in Section 4. Finally, Section 5 contains the conclusion.

where s represents the delineation points, $\tilde{I}$ denotes the testing image, and □ is the training dataset. They tried to find the parameter s which maximizes the probability function $p\ s\ \tilde{I},\quad$. They used 400 images from

(| □)

12 patients to train their network and 50 images to test this network. They obtained an average Hausdorff distance of 0.91 and an average mean absolute distance of 0.86 on the Williams index.

In 2017, Smistad et al. [4] trained a U-Net CNN network [14] and successfully segmented the left ventricle using 2D ultrasound images. However, the issue with this work is its use of the output of a state-of-the-art deformable model segmentation method to train their network. Those models highly depend on the echocardiogram intensities, making it vulnerable in presence of shadows and dropouts. The results showed that the network obtained a Dice Similarity Index (DSI)score of 0.87.

Pinto et al. [5] presented a deep learning method for brain tumor segmentation. In general, their convolutional neural network design is simple: it consists of nine spatial convolution layers and two max-pooling layers. Each spatial convolution uses a 3 3 filter with stride one. The tested image dimensions are (4 33 33), with width and height being 33 pixels each and depth having 4 channels, which will increase to reach 128 activation maps in the 7th layer. However, this method uses a fully connected layer (FC layer) which increases the number of used parameters and the FLOP count. As a result, it slightly increases the required training time. Moreover, having an FC layer at the end of the network will tie the user into using a specific image size (4

33 33) to obtain the segmentation results.

Azarmehr et al. [17] tested the output of three CNN segmentation models (U-Net, SegNet and fully convolutional DenseNets or

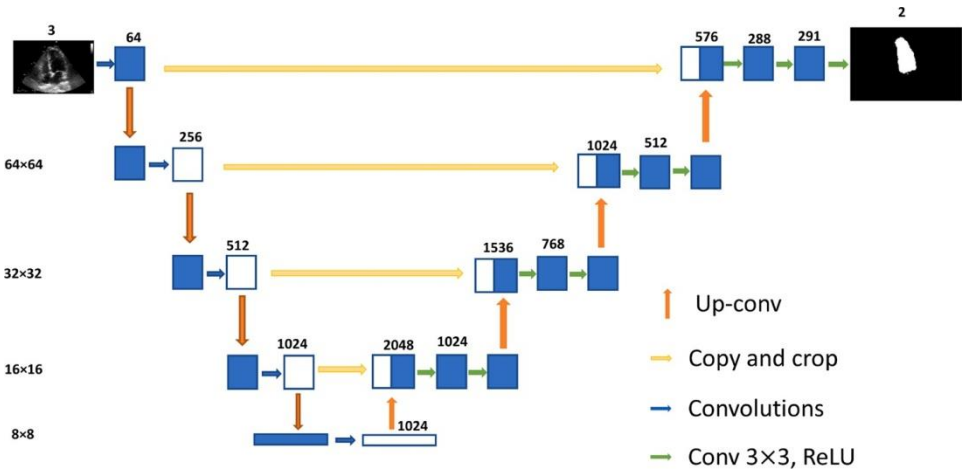

Fig. 1. Model architecture, each blue square block represents a block. To simplify the design, only two blocks are depicted (first and last). The number of channels are above the blocks.

FC-DenseNet). Those models were trained using 992 echocardiograms. The U-Net model outperformed the other models and produced an average DSI of 0.93 ± 0.04.

Hu et al. [18] presented a model based on Bilateral Segmentation Network (BiSeNet) to segment pediatric echocardiograms in 4 chamber view. The proposed model consists of two paths, one path captures low-level features, and the second path captures high-level context features. Also, they used a feature fusion module to fuse the features from those two paths. The model produced DSI of 0.932 when segments the left ventricle, and 0.908 for the left atrium segmentation.

Dong et al. [19] used a model based on voxel-to-voxel conditional generative adversarial nets (cGAN) and the atlas segmentation procedure (VoxelAtlasGAN). This model is based on cGAN which requires generator to produce a segmentation result based on a prior knowledge presented by an atlas. Also, GAN needs a discriminator to discriminate between the resulted generator segmentation and the ground truth. The calculated average DSI of the produced results is equal to 0.953 ± 0.019.

In 2018 Veni et al. [20] used a combination of a trained fully convolutional network and level set to produce a segmentation of LV chamber. The convolutional network is used to produce a segmentation of the LV, which is considered as a prior shape. This shape is then used by the level-set to converge to its final shape. The model produced an average DSI of 0.86 ± 0.06.

Jafari et al. [21] used an approach similar to [19] by using a generative model that maps the masks to their corresponding apical four chamber echocardiograms. The generator is then used as a discriminator to improve the U-Net segmentation results. The propose model produced an average DSI of 93.0 ± 3.9 for ES frames and 94.1 ± 3.3 for ED frames.

There are other approaches which use a deep neural network to track cardiac boundaries [7,12]. However, they are very slow, taking about 20 s to process one frame. Gao et al. [13] reported another deep learning technique which classifies the echocardiography viewpoint. In their training set, they used over 432 videos divided into eight different views such as A2C, A3C, and A4C. In this approach, they used different CNNs. The first CNN (Spatial CNN) is trained using the echo video while the other one received the acceleration using optical flow. With their approach, the process might require weeks to train the classifier; however, unlike the training stage, the system could provide results in real-time.

Oktay et al. [6] utilized CNNs to segment 3D LV structures using an anatomically constrained neural network (ACNN). The segmentation output is constrained to fit a non-linear compact representation of the underlying anatomy derived from an auto-encoder network, which makes it similar to the 3D U-Net [22]. In their work [6], training objective function was defined using a linear combination of cross-entropy (E_r), shape regularisation loss (L_h), and weight decay terms as follows

$$L_h = \| f(\phi(r);\theta) - f(y;\theta) \|_2^2,$$

$$\min_\theta \left(E_r(\phi(r;\theta) - y) + \lambda_1 L_h + \frac{\lambda_2}{2} \| w \|_2^2 \right) \quad (2)$$

Here y represents the labels, r denotes the observed intensity, w corresponds to weights of the convolution filter, and λ_1, λ_2 are the weights of shape regularization loss and weight decay terms used in thetraining. Also, ϕ is a mapping function $\phi : x \longrightarrow y$, and θ denotes the model parameters. We will adopt a similar objective function in our work. The second term in (2) is to ensure the generated segmentation has a low dimensional space as the ground truth labels [6]. In their experiments, they used 3D ultrasound images provided by the CETUS dataset to assess the network, obtaining an average DSI of 0.912.

It is crucial in medical applications to be efficient and reliable. Echocardiographic segmentation methods have to consider a number of issues to provide accurate results. For instance, it is of a great impor- tance to consider occlusions to accurately segment the LV, an issue which is not considered in some other approaches [13]. Also, an auto- matic and real-time segmentation method of the endocardium borders is beneficial for interventional procedures and intensive care unit appli- cations with monitoring requirements. Therefore, it is crucial to have a short inference time compatible with common video rate of 24 frames per second.

3. Proposed approach

This method is based on deep learning algorithms used to segment the LV_{Endo} from cardiographic images. Since it is challenging to obtain a large number of labeled US images to train a CNN network, some data augmentation procedures, such as flipping and scaling the images in the dataset, are used. CNNs provide an estimation for the intensity seg- mentation labels by labeling each pixel in an image and independently taking the surrounding pixels into account. This is accomplished by passing the echocardiogram through sequential convolution layers with a number of filters. Each layer l 1, L consists of f_c channels. Each channel represents a group of neurons that identifies a particular pattern.

Let $y_s = y_{i\ i}$ be a label container which represents different tissue types with y_i 1, 2 which represent foreground and background. In addition, let r r_i $\mathbb{R}, i$ be the captured echocardiogram and be
the total number of the training data. The main purpose of image seg- mentation is to estimate y_s of the captured echocardiogram r. CNN

$\in [\quad]$

$\{\ \}$
$\in \square$
$\in \mathcal{L}\{\quad\}$
$= \{\ \in \quad \in \square\}$ $\square$

segmentation models are performed by training a discriminative function to model a conditional probability distribution $P(y_s|r)$. The evaluation of class densities $P(y_s|r)$ is achieved by assigning a probability to each r_i indicating the pixel's class, generating two sets of class channels f_c that are obtained through sequential convolution layers. The decision for class labels is computed using pixel-wise softmax as in

$$p_c(i) = \frac{e^{(f_c(i))}}{\sum_{c=1}^{C} e^{(f_j(i))}}, \quad (3)$$

where $f_c(i)$ denotes the activation in feature channel c at the pixel position i, and C is the total number of feature channels. Then cross-entropy is defined as

$$E = \sum_{c=1}^{C} \sum_{i \in \Omega} \log\left(\frac{e^{(f_c(i))}}{\sum_{j=1}^{C} e^{(f_j(i))}}\right), \quad (4)$$

and applied to the extracted class activation maps. Similar to U-Net, the mapping procedure between intensities and labels $\phi(r): r \rightarrow \mathcal{L}$ is done by optimizing the average cross-entropy loss of each class $D_r = \sum_{c=1}^{C} E(r, c)$ using stochastic gradient descent.

As mentioned earlier, our network has an analysis (encoder) and a synthesis (decoder) path, each with four resolution steps. In the analysis path, each layer contains several blocks: each block has three convolutions, and three batch-normalizations-and-ReLU-activations. The blocks in the first layer convolve the activation maps with a 1×1 convolution layer to preserve the activation maps' dimensions. The second convolution uses 3×3 kernels and a stride of one. The activation maps' dimensions are maintained by applying padding of one. The third convolution is similar to the first convolution, except for the number of produced activation maps. The number of activation maps differs from one layer to another. For instance, the third and the fourth layers will generate 512 and 1024 activation maps, respectively as in Fig. 1. Furthermore, there is an extra convolution which happens precisely after the first block to reduce the activation map numbers and make it equal to that of the first convolution.

It should be noted that the shortcut in the U-Net are taking place in two directions: horizontally between the decoder and encoder, and vertically between blocks.

Convolutional neural network

CNN has been adopted in several fields, achieving some breakthrough results [23,24]. The CNN layers consist of convolving an image with a number of filters that have the same size within the convolutional layer. This will result in activation maps in which each of their elements is connected to the previous layer through the filter's weights. Also, the number of produced feature maps is equal to the number of applied filters.

Numerous neural network designs strive to obtain the best accuracy and less training time. Some of those networks use a fully connected (FC) layer at the decision-making layers. However, some scholars neglect this part in their design, which results in the reduction of the number of parameters. They may also accommodate more layers to capture further details.

However, the problem with going deeper is having the gradients become infinitesimal because of the backpropagation procedures during the training process. This will result in having the network accuracy drop significantly. However, several approaches have overcome this issue, such as GoogleNet [25] and ResNet [15], which can reach 1200 layers and deliver meaningful improvements [26]. As explained in Section 3.4 we will propose a modified ResNet to accommodate a sufficient number of layers and capture further details without sacrificing accuracy.

Table 1
The followed procedure of multi-stage training for different architectures. (Time measured using PC with NVIDIA TITAN V GPU.)

Architectures	Order	Image size	epoch (min)	Number of epochs	Accuracy
ResNet18 + Unet	1	64×64	0.19	42	89%
ResNet18 + Unet	2	128×128	0.55	70	90.3%
ResNet18 + Unet	3	256×256	3.05	88	91.1%
ResNet32 + Unet	1	64×64	0.22	65	89%
ResNet32 + Unet	2	128×128	0.59	90	90.3%
ResNet32 + Unet	3	256×256	3.020	110	93.32%
ResNet50 + Unet	1	64×64	1.34	60	92%
ResNet50 + Unet	2	128×128	5.05	95	93.42%
ResNet50 + Unet	3	256×256	25.43	120	95.28%
ResU	1	64×64	1.37	70	92.4%

ResU	2	128 × 128	5.12	110	94.13%
ResU	3	256 × 256	26.05	150	97.5%
ResU	–	256 × 256	26.05	192	97.5%

Training details of the proposed model

In this section, we discuss the details of data augmentation used intraining and the training steps. The model is designed to segment 256 256 echocardiographic images. Also, our network does not contain anyFC layer and, hence, enables the use of images with different sizes.Therefore, starting with lower image sizes for training (64 64, 128 128), before reaching 256 256 images, would expedite the trainingprocess where it reduced the required training process by 7 h. Table 1shows the number of required epochs for each architecture to reach theprovided accuracy during a training stage. The last row shows the directintroduction of 256 256 images for training ResU, resulting in 5001.6min for training. The previous three rows exhibit gradual training ofResU using different image sizes, resulting in a total training time of4566.6 min. The reason behind this improvement is the multi-stagegradual training process, which is first carried on 64 64 images, fol-lowed by 128 128 images, before the final stage of training involving
256 256 images. This will result in having the final stage of training to start with a higher Dice accuracy compared to start of training from scratch when applied on 256 256 images. The maximum image size is determined by the computing system capability.

The CAMUS dataset comes with three different image qualities (good, medium, and poor). Leclerc et al. [10] trained their model using good and medium echocardiograms combined, excluding poor quality images, arguing that the network will treat the poor quality data as unimportant entries. In our work, on the other hand, we first exploited the good images then sequentially used the medium images and lastly we did not ignore the images with poor quality. While Leclerc et al. [10]argued that the network treat the poor quality data as unimportant entries, we argue that multi stage training will enable learning of useful details in poor quality images that might be lost otherwise. Therefore, we designed a path of training to force the network to adapt to the new entries of poor quality echocardiograms Fig. 2. The model was first trained using 64 64 good quality images. Starting with a small dataset size is very beneficial, because the processing power and required re- sources are not substantial, making the training process fast compared with the following steps. After training the model using 64 64 good quality echocardiograms, we retrained the model using 128 128 goodquality images. While the outcome was improved, higher demand was
placed on memory and processor, making the training slow. Subse-quently, we retrained the model using 256 × 256 good quality

International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 32, Issue 01) and (Publishing Month: July 2016)
(An Indexed, Referred and Impact Factor Journal)
ISSN: 2319-6564
www.ijesonline.com

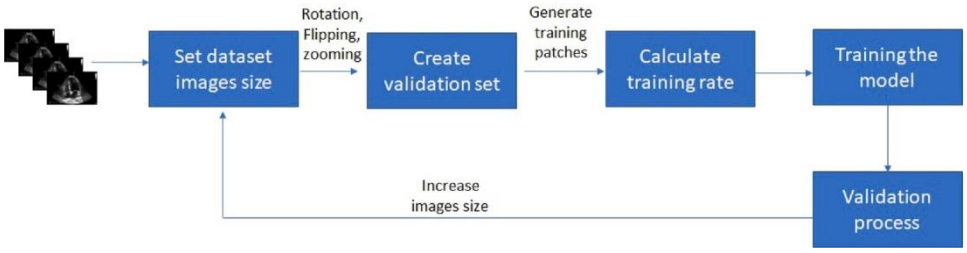

Fig. 2. A flow chart of training and validating process.

echocardiograms. However, due to the limited GPU memory, by using 256 ×256 images, we end up using lower number of batches for training. After this stage of training using good quality echocardiograms, a second training stage was carried out using medium quality echocardiograms with a size of 256×256. After several training epochs, the model started to produce a good Dice similarity index. The same procedures were followed on the poor quality echocardiograms. This training step made use of previously obtained information from both high and medium quality images and updated them using the poor quality images.

Moreover, in deep learning care should be taken during training to ensure generalization capability for the model, avoiding overfit network. Therefore, the training dataset was artificially augmented using affine transformations. The applied transforms are adopted and used as follows:

(1) A random horizontal translation with a probability of 0.5 is applied to improve the network output in the presence of shadows and drop-outs;
(2) A random rotation between−10 and 10 with a probability of 0.75 to avoid adapting the network to a specific orientation;
(3) A random scaling; and
(4) Random contrast.

The probability values selected for this process are complaint with the values suggested by Fastai. These transformations are applied to the training dataset every time we trained our network. Next, we extracted 20% of the training set and kept them as a validation set to provide an unbiased evaluation of the model on the training dataset.

Implementation details

Python language and Fastai library were used to implement the proposed method. All experiments were conducted within Linux OS running Ubuntu 16.04. The training process was carried out on a PC with Intel® i7-8700K CPU @3.70 GHz and 16 GB memory, and NVIDIA TITAN V GPU. To deal with different echocardiogram qualities and representations, the echocardiograms were normalized by the mean value and standard deviation for training purpose. A cross-entropy loss function was utilized with weight decay 1×10^{-4} [15]. Moreover, a batch size that varied from 2 to 8, depending on the echocardiogram dimensions. Also, a variable learning rate (varying between 1×10^{-4} and 1×10^{-6}) based on the cyclic learning rate approach was utilized. The limits of learning rate was selected based on trail and error to provide enhanced accuracy. Furthermore, Adam optimizer was integrated to update the network weights.

Segmentation of LV endocardium

This study aims to provide a reliable and efficient cardiac boundary segmentation method. The common issues that arise in echocardiography images, especially the dropout, make cardiac boundary segmentation difficult for many of the previous segmentation and tracking

methods such as optical flow or block matching. Therefore, in methods that do not use deep learning, scholars tend to combine two techniques. For example Leung et al. [27] integrated affine transformation with optical flow to overcome this issue.

Moreover, the process of validating an automated border detection method in medical images is not an easy task to deal with, especially when several factors (such as image quality and patient data) control theresults.

Our proposed architecture uses U-Net [14] with some modifications on the encoder side. Two factors are considered here: going deeper and improving the segmentation results. With the utilization of the going deeper concept presented in ResNet [15], the network is designed to go deeper in layers without facing a vanishing gradient problem.

Therefore, a hybrid network, namely ResU, was designed by modifying ResNet-50 network and using it as an encoder for the U-Net. The proposed model makes use of echocardiogram information efficiently by strengthening feature propagation throughout the layers, enabling it to preserve as much information as possible from the previous layers. Thismodel was built based on our training strategy in Section 3.2. Also, we used an Adam optimizer in our model and minimized the following costfunction (5):

$$min_\theta \left(E_r(\phi(r;\theta), y \quad \lambda \quad \mathrm{w}\|^2 \right), \qquad (5) \qquad) + \frac{\|}{2} \quad _2$$

where λ is the weight decay term and E_r is defined in Eq. (4).

ResNet50 as encoder

ResNet is best known as a recognition and classification network which provides a very high accuracy rate, whereas U-Net is known as a segmentation network designed to segment biomedical images. Setting a ResNet as an encoder in the U-Net will help improving the classification accuracy of the segmentation network as the number of used parameters is increased, and as a result more spatial information is preserved. Moreover, using a pre-trained ResNet will reduce the required training time and enhance the quality of the segmentation results.

ResNet has several architectures with different number of layers. For instance, ResNet-34 and ResNet-50 are two known ResNet variations, and the difference between these two architectures is in the number of convolutions and batch normalization layers. Unlike other approaches which use max-pooling to reduce the feature maps size, ResNet reduces the feature maps dimensions during spatial convolution operations. Both ResNet-34 and ResNet-50 architectures were tested and ResNet-50 provided better results.

ResU

Any information could be of benefit to segment poorly captured echocardiograms. Therefore, we designed our model based on this concept to provide better segmentation, and to achieve this goal, a modified version of a ResNet is used. ResNet is a very sophisticated network which is used in classification and detection, but it was not created to segment echocardiograms. Each layer in ResNet has several blocks and each block receives a summation of two entries: first entry is the output from the previous layer; with the second one being the output

International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 32, Issue 01) and (Publishing Month: July 2016)
(An Indexed, Referred and Impact Factor Journal)
ISSN: 2319-6564
www.ijesonline.com

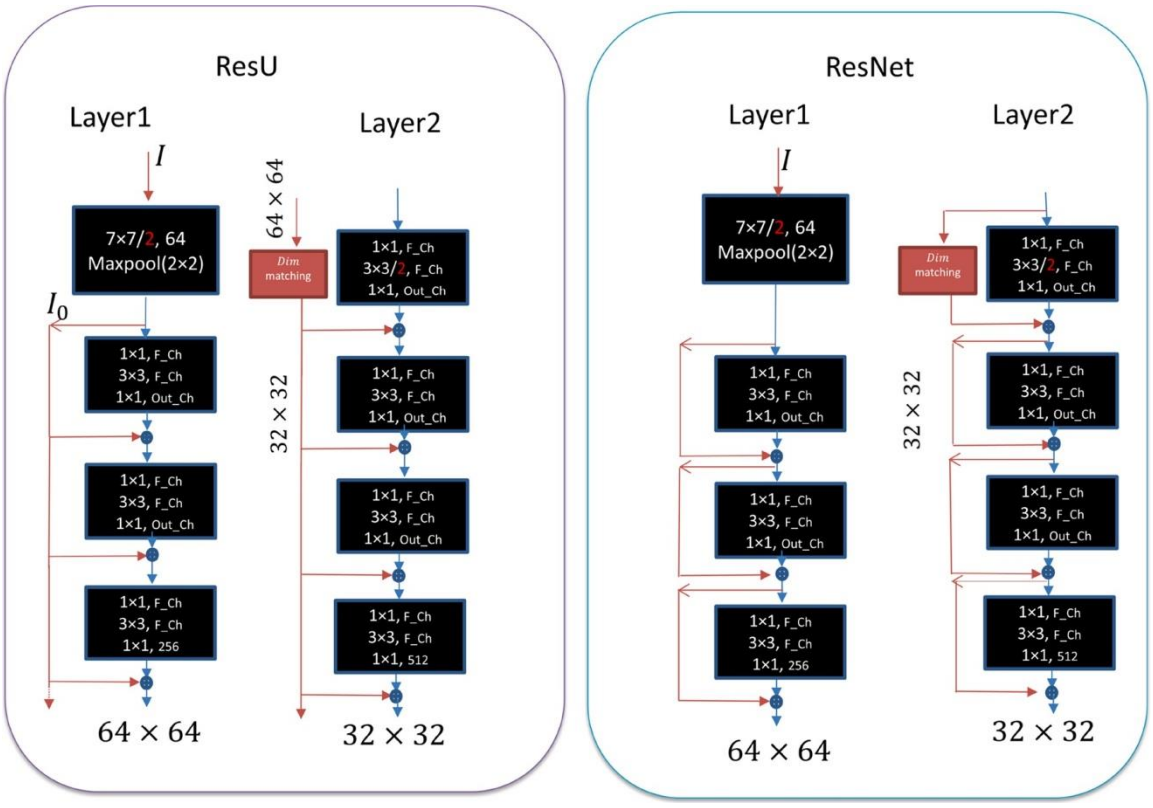

Fig. 3. The first two layers of the proposed ResU (left) and ResNet (right).

from the second to the previous layer. It is crucial to propagate and strengthen the features through the model to enhance echocardiogram information utilization. Accordingly, we spread the originated data from the previous layer to each block in the current layer. ResNet uses the output from the second previous block as an input in the summation operation. However, unlike ResNet, to improve data preservation, our model uses the convolved input data and sums it with each block output, as depicted in Fig. 3.

Suppose a single image I is carried through a convolutional network. The network contains several layers and each layer has a number of blocks which apply a non-linear transformation $F_l(\cdot)$, where l denotes the block indexes. The transformation $F_l(\cdot)$ is composed of operations such as rectified linear units (ReLU), batch normalization (BN), and convolution (Conv). We express the input of the first block in each layer as I_0 and the output of the lth layer as I_l.

Unlike a traditional convolutional network, ResNet adds a connection that holds the input matrix which is summed with the output of the non-linear transformation.

$$I_l = F(I_{l-1}) + I_{l-1} \tag{6}$$

Whereas in our proposed architecture, we keep I_0 to be propagated to each block as in Fig. 3. That is

$$I_l = F(I_{l-1}) + I_0. \tag{7}$$

To elaborate on the difference between ResU and ResNet and demonstrate the advantages of ResU over ResNet, we suggest the following experiment. We feed the system with a zero mean unit variance Gaussian noise. Fig. 4 shows the results. Fig. 4 shows the results of this experiment. Fig. 4a illustrates auto-correlation average of the output of 60 trials for the proposed ResU in Fig. 3. Fig. 4b is the averaged auto-correlation of 60 trials for ResNet. Both Figs. 4a and b show the averaged auto-correlation of the output of the second layer. Also, Fig. 4c demonstrates the averaged auto-correlation on the whole encoder part in

ResU, where Fig. 4d shows the averaged auto-correlation of the output of the entire encoder in ResNet. As expected, not only the output vari- ance of ResU is less than that of ResNet, but also the energy of auto- correlation in ResU is less than that of ResNet. This is due to the fact that noise in ResNet will propagate with a higher correlation than in ResU as the input and the output of each block are being added in ResNet. Therefore, there is a higher chance of a higher correlation be- tween these noises in ResNet. On the other hand, in ResU the output is being added to the input of the layer which can be further from the input of that block which has less correlation with that output. This experi- ment illustrates what happens to the additive noise of the image in both ResNet and ResU and indicates efficiency of ResU over ResNet in dealing with this noise.

As in a traditional U-Net, our proposed model decreases the feature maps dimensions at each layer. It receives an image with a resolution of
256 256 and the last layer in the encoder will have 8 8 feature maps. Also, we trained our model using learning rates calculated based on the cyclic learning rate proposed by Smith et al. [28]. In addition, to conserve memory usage, we changed batch size based on the used image dimensions. For instance, a batch size of two is used when 256 256 image dimensions were utilized in training. Table 2 provides the used normalization, batch size and the learning rate in the proposed architecture.

The first layer in ResU will generate a 64 activation map. The next layer will receive the generated activation map and convolve it, creating a 128 activation map. This process will continue through the remaining layers, until the activation map reaches 2048.

Each layer results are kept to be used during the deconvolution part (decoder). This can be expressed as a set of functions, each of which depends on the output of the previous function as follows

$$F_e = F_5(F_4(F_3(F_2(F_1(I_1, w_1), w_2), w_3), w_4), w_5). \quad (8)$$

Similar to the encoder, the decoder has several sub-parts. It will receive the calculated activation maps during the encoding process and

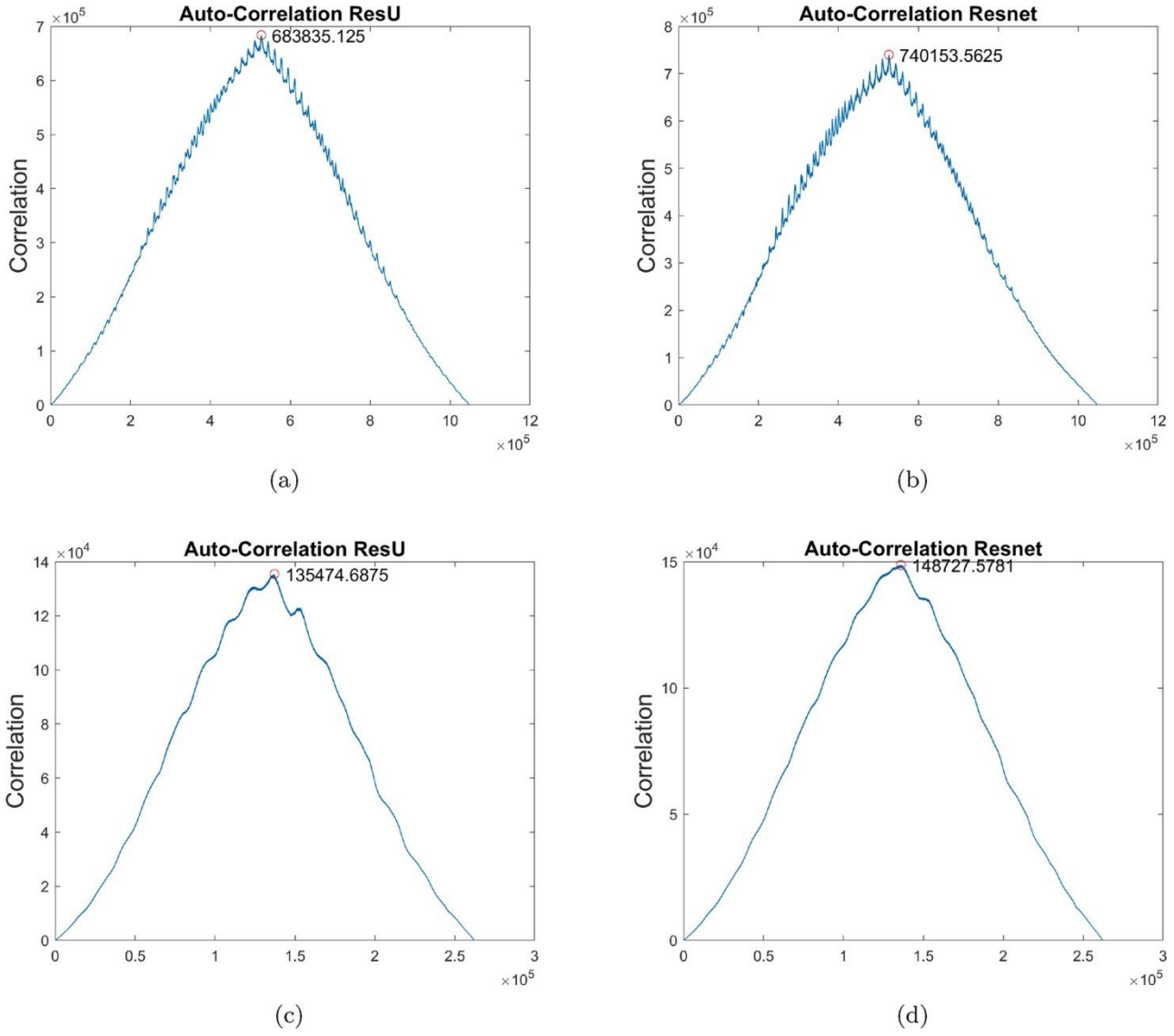

Fig. 4. The averaged auto-correlation of output to a white a unit variance Gaussian noise (representing additive noise), averaged over 60 trials. (a, c): ResU results, (b, d): ResNet results. (a, b) are the output of the second layer and (c, d) show the output of the last layer.

Table 2
The characteristics of U-Net 2, ResNet-U-Net and the proposed approach.

Architectures	Lowest Res	Normalization	Batch size	Learning rate
U-Net 2 [10]	8 × 8	BatchNorm	8	1e—3
ResNet-U-Net	8 × 8	BatchNorm	4	1e—4 To 1e—5
Proposed	8 × 8	BatchNorm	2-8	1e—4 To 1e—6

Table 3
A comparison of LV segmentation with no distinction of the chamber current phase (mean ± standard deviation).

Methods	DSI	HD
[17]	0.93 ±0.04	4.52 ±0.9
[18]	0.93 NA	NA
[19]	0.953 ±0.03	NA
Proposed	**0.975** ±0.0109	**2.562** ±0.7242

merge it with the de-convolved activation map (depth-wise). Also, unlike the encoder, the decoder will reduce the number of channels by half. For instance, the first subpart in the decoder will receive an activation map of 1024-depth. Furthermore, it will combine it with generated activation maps from the previous encoder layer, which results in an activation map with a depth of 2048 which is reduced to 1024 after a convolution step. Moreover, decoders increase the activation map dimensions, which is the opposite of what happens in the encoder part. The reduction of the depth number and the increase in the activation map sizes will continue until it reaches a depth of two, and the dimension of 256 ×256 which is a mask representing background and foreground.

4. Simulations

In our simulations, ResU was tested it on four different groups of endocardiograms. The first group of echocardiograms contains LV_{Endo} in end-diastole (ED) phase. The second group includes echocardiograms of LV_{Endo} in end-systolic (ES) phase. The third group consists of only poor quality echocardiograms. The final group consists of both ES and ED phases in echocardiograms with good and medium quality. All validation groups consist of both two-chamber and four-chamber echocardiograms.

Dataset

The CAMUS [10] dataset was used to train and test the segmentation network. This dataset consists of clinical exams from 500 patients. Also, it enforces clinical realism by including cases that are difficult to trace, cardiac walls that are invisible, and a wide variate of acquisition settings. The dataset provides both ED and ES for two chambers and four chambers with extra information about the patients' ejection fraction (EF) and image quality, which comes in three different categories (good, medium and poor). The main problem for labeling 2D echocardiograms is the artifacts which accompany the echocardiograms, such as shadows and low contrast. For medical images, labeling is usually expensive to acquire because both professionals and time are needed to provide reliable annotations. Three cardiologists (O_1, O_2, O_3) participated in the annotation of the CAMUS dataset. The first cardiologist defined a consistent segmentation protocol which was followed by the other cardiologists [10].

Also, due to the lack of a reliable ECG, the cardiologists did not follow the recommendations of both the American Society of Echocardiography and the European Association of Cardiovascular Imaging in defining both ES and ED frames. The method of selecting ED and ES in this dataset was by observing the LV dimensions. Moreover, CAMUS was divided into 10 parts, with each part containing images from 50 patients. We used one part for testing and the remaining (9 parts) for training ResU.

Evaluation metrics

To measure the accuracy of segmenting LV_{Endo}, three different metrics are used, namely, Dice Similarity Index (DSI) [29], the 2D Hausdorff distance (HD), and the mean absolute distance (MAD) are used to evaluate the segmentation accuracy.

Let $U = \{u_1, u_2, ..., u_m\}$ be the predicted area, and $R = \{r_1, r_2, ..., r_m\}$ be the reference area, S_U represent the predicted region enclosed by U, and S_R denote the reference region enclosed by R.

DSI measures the overlap between the regions from the manual and automatic segmentation techniques. That is

$$\mathrm{DSI} = 2\,\frac{(S_U \cap S_R)}{(S_U + S_R)}, \tag{9}$$

where S_U, S_R are predicted surface and the reference area, respectively.

Also HD and MAD are defined as follows:

$$\mathrm{HD} = max\ \ max_i\{d(u_i, R)\} + max_j\{d(r_i, U)\}\Big), \tag{10}$$

Table 4
Segmentation accuracy of the 3 evaluated methods and the three cardiologists. ED: end diastole, ES: end systole, DSI: Dice similarity index, MAD: mean absolute

distance, HD: Hausdorff distance, APT: automated processing time (mean ± standard deviation). [6] (GPU: Nvidia GTX-1080), [10] (GPU: Nvidia Tesla M60).

Methods	ED			ES			APT
	DSI	MAD	HD	DSI	MAD	HD	
O_1 vs. O_2	0.919 ±0.033	2.2 ±0.9	6.0 ±2.0	0.873 ±0.060	2.7 ±1.2	6.6 ±2.4	N/A
O_1 vs. O_3	0.886 ±0.050	3.3 ±1.5	8.2 ±2.5	0.823 ±0.091	4.0 ±2.0	8.8 ±3.5	N/A
O_2 vs. O_3	0.921 ±0.037	2.3 ±1.2	6.3 ±2.5	0.888 ±0.058	2.6 ±1.3	6.9 ±2.9	N/A
[6]	0.932 ±0.034	1.7 ±0.9	5.8 ±3.1	0.903 ±0.0.059	1.9 ±1.1	6.0 ±3.9	0.06 s
[10]	0.939 ±0.043	1.6 ±1.3	5.3 ±3.6	0.916 ±0.061	1.6 ±1.6	5.5 ±5.5	0.09 s ±0.03
Proposed	**0.975** ±0.0086	**0.006** ±0.0032	**2.846** ±0.8014	**0.972** ±0.0106	**0.0057** ±0.0018	**2.61** ±0.6723	**0.05 s** ±0.004

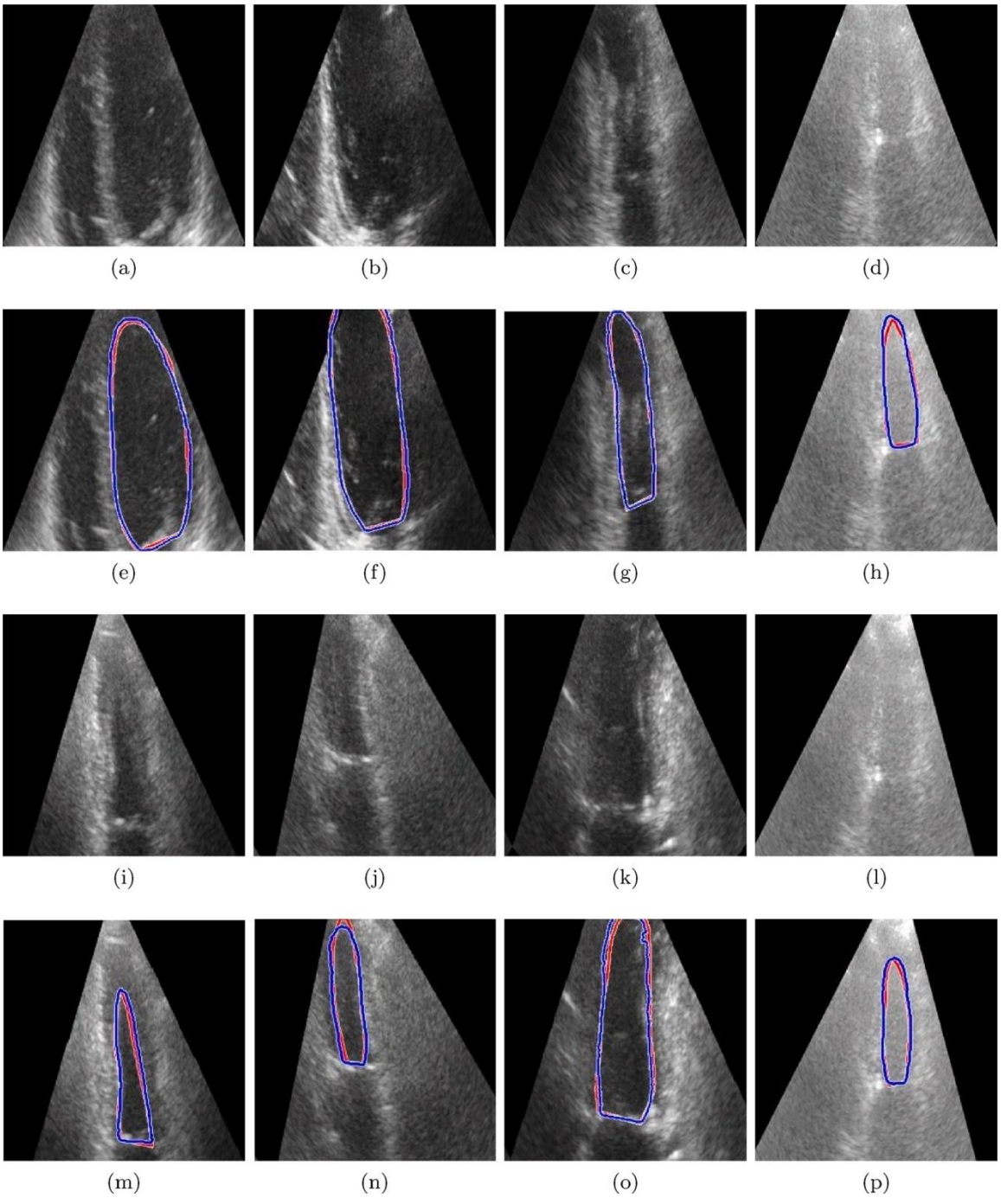

Fig. 5. The first and the third rows are the original images, the second and the fourth rows contains the performance of our model (blue) and manual boundary segmentations (red). (d and h) are poor quality images, and the rest are medium quality images.

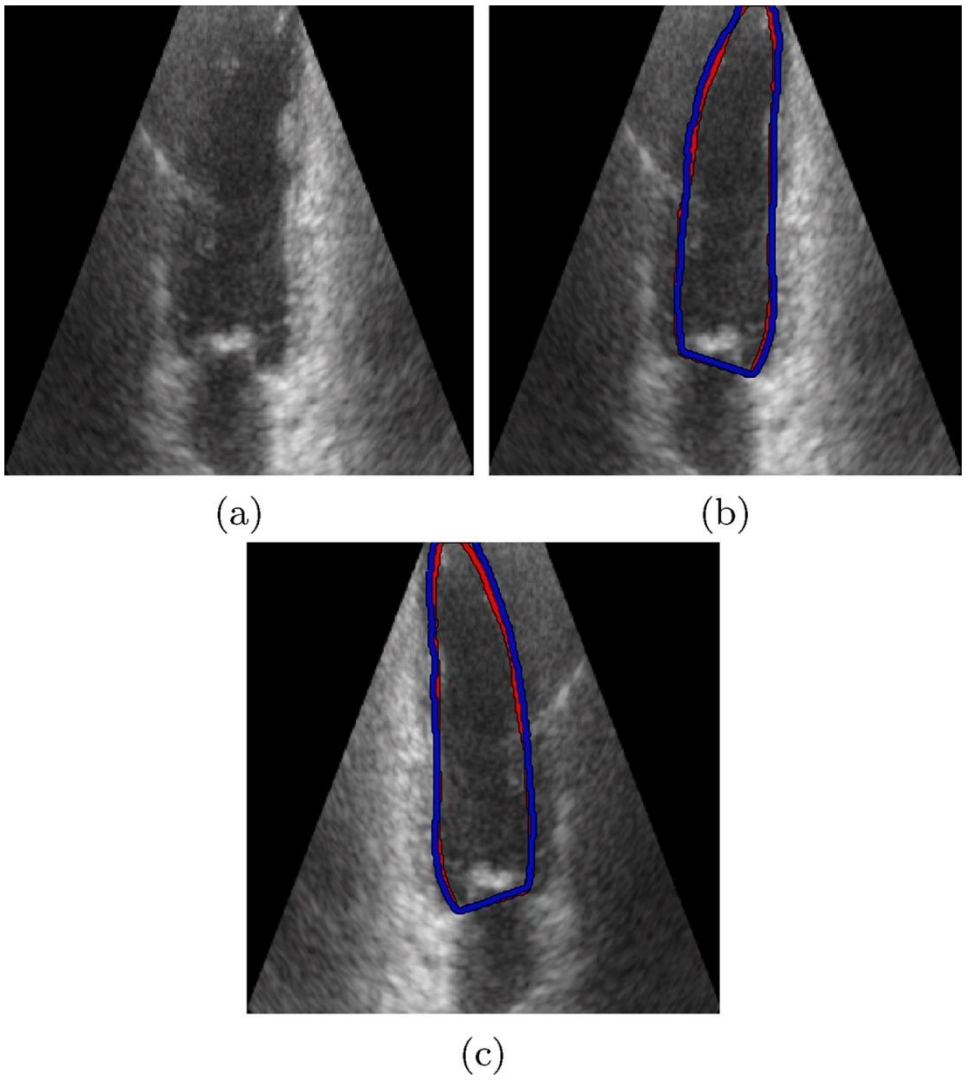

Fig. 6. Our proposed model gives good results even with flipping the testing images (blue) and manual boundary segmentations (red).

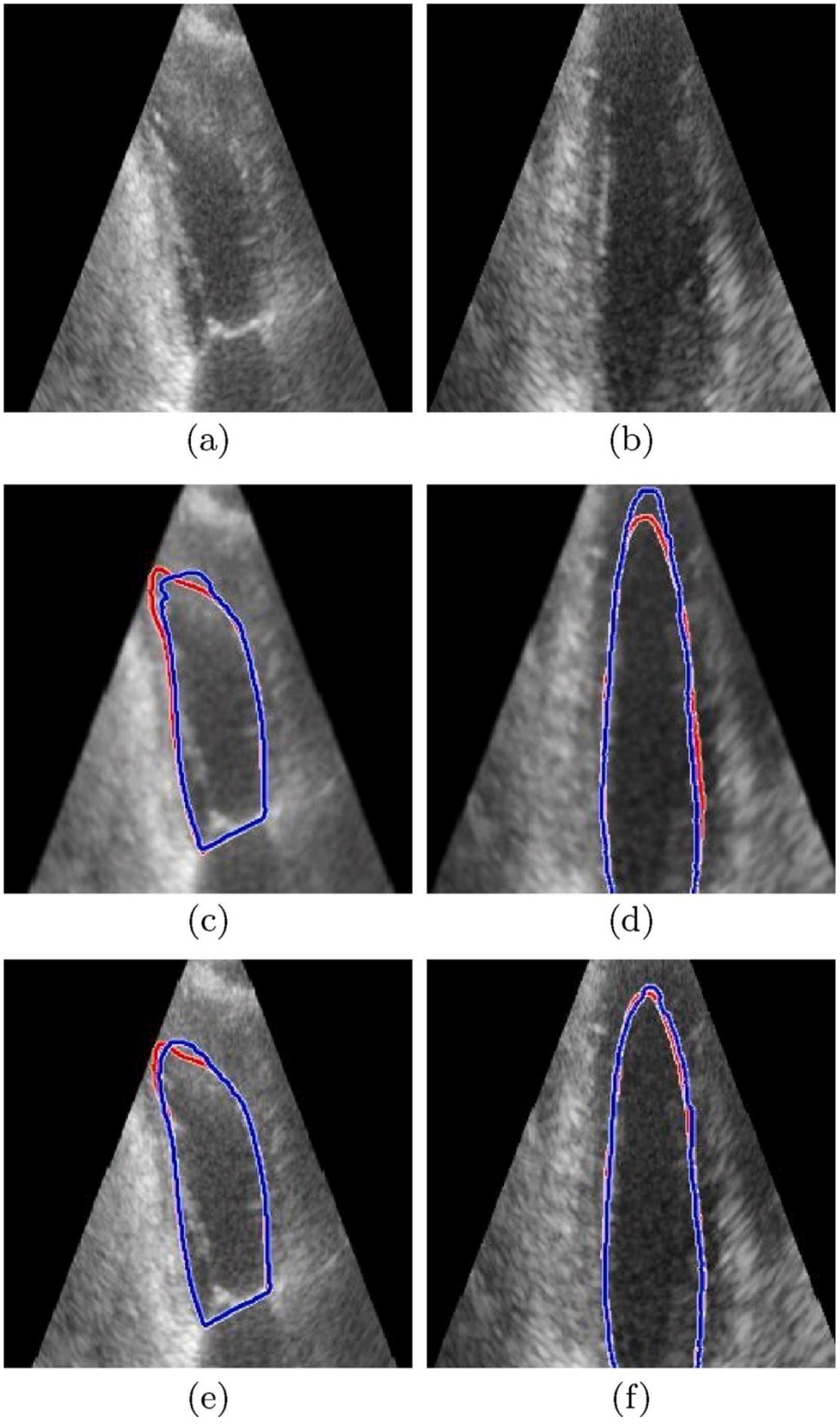

Fig. 7. The effect of blurring echocardiograms on our model predictions (c and d are the blurred versions of a and b).

Table 5
The effect of different selected transformations on the model prediction.

Transformation	DSI	MAD	HD
Rotation	0.967	0.0092	3.0340
	±0.0214	±0.0048	±0.9038
Horizontal flipping	0.9793	0.0058	2.5306
	±0.0083	±0.0023	±0.7419
Scaling	0.9626	0.0155	3.3873
	±0.0219	±0.0073	±1.0172
No transformation	0.9799	0.0056	2.5098
	±0.0077	±0.0021	±0.7128

Table 6
The effect of using Gaussian filter with different sigmas on the model prediction.

σ	DSI	MAD	HD
0.5	0.9699	0.2061	2.6848
	±0.0145	±0.0856	±0.7044
0.7	0.9667	0.2035	2.7652
	±0.0162	±0.0858	±0.7338
0.9	0.9574	0.2025	2.8940
	±0.0525	±0.0878	±0.7718
1.2	0.9543	0.2041	3.0412
	±0.0289	±0.0870	±0.7547
1.5	0.8981	0.2017	3.5122
	±0.1273	±0.0840	±0.9386

$$MAD = \frac{1}{2}\left\{\frac{1}{n}\sum_{i=1}^{n} d(u_i, R) + \frac{1}{n}\sum_{j=1}^{n} d\ r_j, U\right)\right\}, \quad (11)$$

where $d(u_i, R) = min_j \parallel r_j - u_i \parallel$.

MAD is used to represent the global disagreement between two contours with HD suggesting how far these contours are from each other.

Empirical results

The purpose of the simulations was to verify the method and compare its performance with the benchmarks established in CAMUS and previous established approaches.

Table 3 provides the accuracy of the methods with no distinction of the chamber current phase, and it shows that ResU provided higher DSI than other deep-learning-based segmentation methods. Table 4 represents the segmentation accuracy calculated from echocardiograms with good, medium, and poor image qualities. Mean and standard deviation values for each metric were obtained by following the same strategy as Leclerc et al. [10]. Fig. 5 depicts the ground truth and the prediction of three samples from the validation dataset. From these results, it is concluded that our proposed method obtains better segmentation scores for both ES and ED. Moreover, our approach is robust to image affine transformations such as scaling, rotation, and horizontal flipping, with example being shown in Fig. 6.

Moreover, the results of our model predictions on blurred and non-blurred echocardiograms are depicted in Fig. 7.

Furthermore, random translations were applied on the same echocardiograms to test the model accuracy on predicting the LV walls, and Table 5 show the ResU results.

For a real-time application, it is crucial to have a short inference time compatible with common video rate of 24 frames per second. The proposed system has inference time of 0.050 0.0046 s for segmenting a single image, using a PC with Intel® i7-8700K CPU @3.70 GHz.

In Section 3.1, we mentioned that during the training process, the deep learning network starts to learn patterns and shapes. First, layers will identify basic shapes, as those shapes become more and more sophisticated, the deeper we go into the network. Hence, this network will identify the object or objects that are included in the training set based on a scoring mechanism. Therefore, a small modification in the object shape will result in a decrease in the decision score. Accordingly, we used scaling and rotation, and other tools in our augmentation process to improve the model's predictability. For instance, dropouts are one of the issues that accompany echocardiograms and, to deal with this issue, we augmented the training data set with random scaling and rotation functions. The use of augmentation resulted in a model that can reliably predict the LV_{Endo} walls.

The model performance tested on five different sets of images. Two sets are grouped based on the cardiac phase (ES, ED), and the rest are based on the echocardiograms quality (good, medium, poor). Also, to test our model robustness, some random affine transformations were applied, such as scaling, rotating, and translation.

Each of those echocardiogram sets was tested two times (with and without affine transformations). The first set contained echocardiograms in ES phase. The set was first tested without applying affine transformations and it produced a DSI of 0.9807 ± 0.001, and after applying affine transformations, it reached a DSI of 0.9747 ± 0.0076. The second set contained the cardiograms in ED phase, and the model generated DSI of 0.9778 ± 0.0035 and 0.9758 ± 0.0086, before and after applying affine transformations, respectively.

The next three sets contained images with various quality. First group had good quality echocardiograms, for which our model produced DSI of 0.9811 ± 0.004, 0.9758 ± 0.0086, before and after affine transformations, respectively. The second group contained medium quality echocardiograms, for which our model generated DSI of 0.9773 ± 0.0062, and 0.9758 ± 0.00106, before and after applying affine transformations, respectively. Thirdly, we tested our model performance on poor quality echocardiograms, and with those images, DSI of 0.9724 ± 0.0102, and 0.9736 ± 0.0098 were reached, before and

after the application of affine transformations, respectively.

In image processing, edges are characterized by sharp changes in intensities. Thus, blurring filters have the undesirable effect on correctlysegmenting the edges. Therefore, we randomly selected a set of echo- cardiograms from the validation dataset and applied a Gaussian filter with σ range between 0.5 and 1.5 to produce the blurring effect on the echocardiograms. Table 6 provides the model segmentation accuracy when Gaussian filter is applied on set of echocardiograms. This drop in accuracy was expected because the blur function averaged the in- tensities in the echocardiograms.

As a result, the edges are less visible than those in the training set, making our model unable to find some parts of the LV_{Endo}.

The training procedure in Section 3, significantly improved the segmentation results. Hence, instead of starting a training stage on pa- rameters from a random state, we used the acquired knowledge from the previous stage to update the parameters with new trained entries. Therefore, we started training the network using images with 64 64 dimensions to initialize the model parameters. This is a pre-training step which initializes the model's weights. Also, as a pre-training stage, we selected 64 64 images to reduce the required training process time. Therefore, each training stage benefited from the previous training stage, and it continued improving the segmentation results as more details became available. This procedure was repeated three times. It produces 0.92 DSI, 0.94 DSI, 0.97 DSI in the first, second and the last training stages, respectively. Also, it includes results pertaining to the three cardiologists who annotated the LV_{Endo} in the CAMUS dataset.

As one can note, our proposed method obtained better DSI scores (mean DSI of 0.975 at ED and 0.972 at ES), HD (mean HD of 2.846 at ED and 2.61 at ES) and MAD (mean MAD of 0.0079 at ED and 0.0057 at ES). These results show the effectiveness of our proposed model. Also, our proposed approach is able to provide a good segmentation results of poor quality echocardiograms. Moreover, ResU is able to segment LV borders even after applying random translations on the echocardiogram

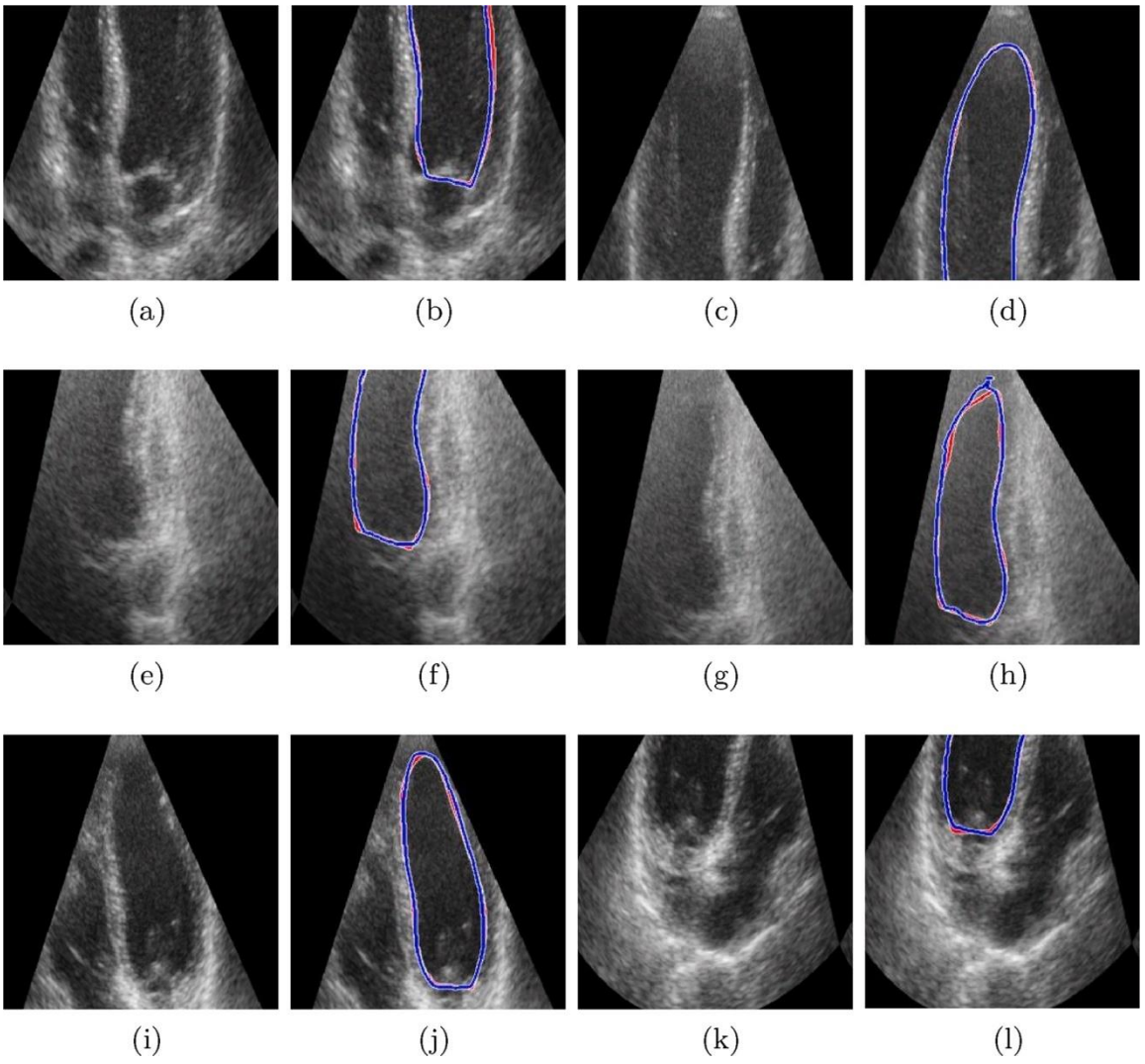

Fig. 8. Random translations are applied on the same echocardiograms to test the model ability to predict the LV walls.

as in Fig. 8.

5. Conclusion

In this work, we introduced a hybrid net, denoted by ResU (using U-Net and a modified version of Res-Net), and an efficient automatic segmentation approach for echocardigraphic images. In addition to propagating and strengthening the features throughout the model, we spread the originated data from the previous layer to each block in the current layer. However, unlike ResNet, our model uses the convolved input data and sums it with each block output. Results show that utilizing the pre-trained network has improved the model output. The model produced a DSI score of 0.97 on the testing set. The segmentation results of this model are very close to the manual annotation of the experts. In the future work, we plan to apply the proposed model to other applications, such as calculating the ED and ES volumes.

Authors' statement

Yasser Ali: conceptualization, methodology, software, writing-original draft preparation, revising. **Farrokh Janabi-Sharifi**: co-supervision, conceptualization, investigation, resources, project administration, funding acquisition, writing – reviewing and editing, revising. **Soosan Beheshti**: co-supervision, investigation, funding acquisition, writing – reviewing and editing, revising.

Acknowledgments

The authors would like to thank Dr. Aleksandar Vakanski of Idaho University for his help in improving the readability of the paper. This work was financially sponsored by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Discovery Grant 2017-06930. The first author was also partially supported through a scholarship from the Libyan Government. Also, we thank Nvidia for their support by supplying the Titan V graphic card used in our experiments.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] V. Tavakoli, N. Bhatia, R.A. Longaker, M.F. Stoddard, A.A. Amini, Tissue Doppler imaging optical flow (TDIOF): a combined B-mode and tissue Doppler approach forcardiac motion estimation in echocardiographic images, IEEE Trans. Biomed. Eng. 61 (8) (2014) 2264–2277.

[2] X. Zhang, M. Günther, A. Bongers, Real-time organ tracking in ultrasound imagingusing active contours and conditional density propagation, in: International Workshop on Medical Imaging and Virtual Reality, Springer, Berlin, Heidelberg, 2010, pp. 286–294.

[3] A. Skalski, P. Turcza, Heart segmentation in echo images, Metrol. Meas. Syst. 18 (2) (2011) 305–314.
[4] E. Smistad, A. Østvik, 2D left ventricle segmentation using deep learning, in: 2017 IEEE International Ultrasonics Symposium (IUS), IEEE, 2017, pp. 1–4.
[5] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251.
[6] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S.A. Cook, A. De Marvao, T. Dawes, D.P. O'Regan, B. Kainz, Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation, IEEE Trans. Med. Imaging 37 (2) (2017) 384–395.
[7] G. Carneiro, J.C. Nascimento, A. Freitas, The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods, IEEE Trans. Image Process. 21 (3) (2011) 968–982.
[8] A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Trans. Med. Imaging 25 (8) (2006) 987–1010.
[9] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models—their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59.
[10] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A.R. Berg, P.M. Jodoin, T. Grenier, C. Lartizien, Deep learning for segmentation using an open large-scale dataset in 2d echocardiography, IEEE Trans. Med. Imaging (2019).
[11] G.B. Bleeker, P. Steendijk, E.R. Holman, C.M. Yu, O.A. Breithardt, T.A. M. Kaandorp, M.J. Schalij, E.E. Van der Wall, P. Nihoyannopoulos, J.J. Bax, Assessing right ventricular function: the role of echocardiography and complementary technologies, Heart 92 (Suppl. 1) (2006) i19–i26.
[12] G. Carneiro, J.C. Nascimento, Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2592–2607.
[13] X. Gao, W. Li, M. Loomes, L. Wang, A fused deep learning architecture for viewpoint classification of echocardiography, Inf. Fusion 36 (2017) 103–113.
[14] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 234–241.
[15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016) 770–777.
[16] D. Barbosa, T. Dietenbeck, B. Heyde, H. Houle, D. Friboulet, J. D'hooge, O. Bernard, Fast and fully automatic 3-d echocardiographic segmentation using B-spline explicit active surfaces: feasibility study and validation in a clinical setting, Ultrasound Med. Biol. 39 (1) (2013) 89–101.
[17] N. Azarmehr, X. Ye, S. Sacchi, J.P. Howard, D.P. Francis, M. Zolgharni, Segmentation of Left Ventricle in 2D echocardiography using deep learning, in: Annual Conference on Medical Image Understanding and Analysis, Springer, Cham, 2019, pp. 497–504.
[18] Y. Hu, L. Guo, B. Lei, M. Mao, Z. Jin, A. Elazab, B. Xia, T. Wang, Fully automatic pediatric echocardiography segmentation using deep convolutional networks based on BiSeNet, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 6561–6564.
[19] S. Dong, G. Luo, K. Wang, S. Cao, A. Mercado, O. Shmuilovich, H. Zhang, S. Li, VoxelAtlasGAN: 3D left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2018, pp. 622–629.
[20] G. Veni, M. Moradi, H. Bulu, G. Narayan, T. Syeda-Mahmood, Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 898–902.
[21] M.H. Jafari, H. Girgis, A.H. Abdi, Z. Liao, M. Pesteie, R. Rohling, K. Gin, T. Tsang, P. Abolmaesumi, Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as prior, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 649–652.
[22] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2016, pp. 424–432.
[23] S. Dieleman, K.W. Willett, J. Dambre, Rotation-invariant convolutional neural networks for galaxy morphology prediction, Mon. Not. R. Astron. Soc. 450 (2) (2015) 1441–1459.
[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012) 1097–1105.
[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 1–9.
[26] G. Huang, Y. Sun, Z. Liu, D. Sedra, K.Q. Weinberger, Deep networks with stochastic depth, in: European Conference on Computer Vision, Springer, Cham, 2016, pp. 646–661.
[27] K.E. Leung, M.G. Danilouchkine, M. van Stralen, N. de Jong, A.F. van der Steen, J. G. Bosch, Left ventricular border tracking using cardiac motion models and optical flow, Ultrasound Med. Biol. 37 (4) (2011) 605–616.
[28] L.N. Smith, Cyclical learning rates for training neural networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2017, pp. 464–472.
[29] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, W. M. Wells III, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports, Acad. Radiol. 11 (2) (2004) 178–189.