LVNet: Lightweight Model for Left Ventricle Segmentation for Short Axis Views in Echocardiographic Imaging

Navchetan Awasthi, *Member, IEEE*, Lars Vermeer, Louis S. Fixsen,
Richard G. P. Lopata, *Senior Member, IEEE*,
and Josien P. W. Pluim, *Fellow, IEEE*

***Abstract*—Lightweight segmentation models are becoming more popular for fast diagnosis on small and low cost medical imaging devices. This study focuses on the segmentation of the left ventricle (LV) in cardiac ultrasound (US) images. A new lightweight model [LV network (LVNet)] is proposed for segmentation, which gives the benefits of requiring fewer parameters but with improved segmentation performance in terms of Dice score (DS). The proposed model is compared with state-of-the-art methods, such as UNet, MiniNetV2, and fully convolutional dense dilated network (FCdDN). The model proposed comes with a post-processing pipeline that further enhances the segmentation results. In general, the training is done directly using the segmentation mask as the output and the US image as the input of the model. A new strategy for segmentation is also introduced in addition to the direct training method used. Compared with the UNet model, an improvement in DS performance as high as 5% for segmentation with papillary (WP) muscles was found, while showcasing an improvement of 18.5% when the papillary muscles are excluded. The model proposed requires only 5% of the memory required by a UNet model. LVNet achieves a better trade-off between the number of parameters and its segmentation performance as compared with other conventional models. The developed codes are available at https://github.com/navchetan-awasthi/Left_Ventricle_Segmentation.**

***Index Terms*—Deep learning (DL), lightweight models, neural networks, segmentation, ultrasound (US) imaging.**

I. Introduction

ULTRASOUND (US) imaging is a cheap and noninvasive technique, which can be easily utilized for assessing the geometry and function of the heart [1]. Segmentation of the left ventricle (LV) of the heart is a key task in the measurement of cardiac function parameters, such as ejection fraction, systolic and diastolic volumes, and myocardium mass, which are all indicators of cardiac health. The segmentation of the LV also aids in regional analysis and surgical planning [2], [3]. Thus, segmentation of LV is an important step in diagnosis of cardiopathies albeit a challenging task because of the limited and inhomogeneous contrast between myocardium and endocardium, which is due to physical artifacts, such as acoustic shadowing, edge dropouts, clutter, speckle noise, and so on. There are various techniques for segmentation of the endocardium, such as classical feature extraction techniques [4], deformable models [5], and active contour methods [6], [7]. These techniques have various limitations: using computationally expensive feature extraction, requiring prior information about the tissue properties, and overall lack of robustness [8]–[10].

In this study, we focus on the segmentation of the LV in images of the short axis views of the heart. Short axis views are clinically relevant and commonly acquired in a standard echo test, and hence, an automatic algorithm for these views will benefit the community. There are two types of segmentation, which we discuss in this study—with papillary (WP) and without papillary (WOP). Identification of the papillary structure is important for various reasons. Segmentation of the LV WP structure helps in accurately and easily determining the LV mass for predicting cardiovascular morbidity [11]–[13] while the segmentation of the region WOP muscle helps in assessing the cardiac output, ejection fraction, and in cardiac strain imaging [12], [14]. Few of the segmentation studies address the importance of including or excluding the papillary muscles while performing segmentation [1], [15]. Manual segmentation is time-consuming and a challenging process because of factors, such as low contrast and poor boundary definition. Therefore, automated segmentation of such structures in US images is of utmost importance for determining the cardiac function in an accurate manner.

Previously, many methods have been developed for LV segmentation, including machine learning-based methods, e.g., structured random forest was used, but it requires manual selection of features [16]. Various deep learning (DL)-based

Manuscript received February 11, 2022; accepted April 17, 2022. Date of publication April 22, 2022; date of current version May 26, 2022. This work was supported in part by the 4TU Precision Medicine Program through the High Tech for a Sustainable Future, a framework commissioned by the four Universities of Technology of the Netherlands. *(Corresponding author: Navchetan Awasthi.)*

Navchetan Awasthi, Lars Vermeer, Louis S. Fixsen, and Richard G. P. Lopata are with the Department of Biomedical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: n.awasthi@tue.nl).

Josien P. W. Pluim is with the Department of Biomedical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands, and also with the Image Sciences Institute, University Medical Center at Utrecht, 3508 GA Utrecht, The Netherlands.

Digital Object Identifier 10.1109/TUFFC.2022.3169684

models were also developed for the same task of LV segmentation [3], [17]. DL techniques have been shown to be effective for a variety of tasks, such as classification [18], [19], segmentation [20], [21], super-resolution [22], [23], reconstruction [24], [25] of images in medical applications, and other areas. Recently, DL architectures, such as convolutional neural networks (CNNs), have shown promise for cardiac image segmentation by improving the segmentation quality and speed [17], [26]. These models are currently being used widely for 2-D and 3-D US cardiac segmentation [27], [28]. Veni *et al.* [29] proposed a DL-based framework utilizing a UNet followed by a level set framework for refining the results for LV segmentation in B-mode echocardiography images. Jafari *et al.* [30] proposed a combination of UNet and bidirectional LSTM for capturing the spatiotemporal information simultaneously for improving the LV segmentation.

In recent years, segmentation models have been optimized in terms of size and memory requirements compared with larger state-of-the-art models, such as the UNet. DL has shown to be a very promising technique for segmentation in medical images; however, standard networks are computationally expensive. Lightweight networks require less computational power and, hence, can be easily deployed on resource-constrained devices. The successful implementation of these models can result in almost real-time segmentation [31]. Thus, lightweight segmentation architectures are an important and interesting field of investigation for implementation of image processing pipelines, such as segmentation tasks. The number of parameters of the network architecture has been reduced substantially while maintaining or even improving segmentation performance [32]–[34]. Lightweight CNNs have proven to form a basis for efficient and accurate semantic segmentation tasks for medical images [35]–[37], such as detecting lung pathologies, abdominal structures, tumors, and heart structures with the advantage of fast and accurate prognosis [1], [36], [38], [39]. In this study, we focus on the segmentation of LV in US images of the short axis views of the heart using lightweight CNNs. A deep model (without overfitting and underfitting) with a higher number of layers is supposed to perform better than a model with a lower number of layers. Still, the focus is shifting toward shallow models, because these have demonstrated to provide similar results as the deep model architectures [40]. As the model becomes more shallow, the number of parameters decreases, which will lead to a lower demand in terms of computing resources, and the costs associated. Thus, lightweight models with fewer parameters are an important step to reduce the dependency on costly computing resources [41], [42]. Hence, lightweight models are highly suitable for implementation in low-cost hardware in low-resource settings where memory usage forms a restriction for implementation of DL models [34]. This will allow faster, point-of-care diagnosis by providing cardiac geometry and function without human intervention.

This study proposes a new lightweight model architecture [LV network (LVNet)], which requires fewer parameters while achieving better performance as compared with other state-of-the-art architectures, i.e., the well-established UNet, and two different lightweight algorithms called MiniNetV2 and fully convolutional dense dilated network (FCdDN) [34], [43]. The performance of UNet and existing lightweight architectures, including MiniNetV2, a state-of-the-art network, and FCdDN, a network proposed by Ouahabi and Taleb-Ahmed, were compared on LV US images. In the original paper, the FCdDN model portrayed promising results for US thyroid images as compared with other lightweight networks, such as E-net [34], [43]. Several modifications are proposed to improve the trade-off between the performance and the amount of parameters. The architecture details of UNet, MiniNetV2, FCdDN, and the proposed model will be given later in Section II-D.

This article makes the following contributions in the field of LV segmentation.

1) Introduction of a new lightweight segmentation network that has a better trade-off between the number of parameters and the segmentation performance.
2) Introducing a different training strategy combined with a post-processing pipeline, which can also be used for other segmentation tasks.
3) Introduction of a parameter reduction technique by incorporating a channel split to extract features, whereas most model architectures perform single channel convolutions.
4) The proposed model is shown to perform favorably as compared with state-of-the-art DL-based techniques.

II. Related Work

A. Related Work

In the previous studies, typical state-of-the-art networks, such as UNet, produce accurate results but are computationally costly because of their complex architectures [33], [44]. Other methods, such as the one proposed by Hsu [1], combine Faster-region-based CNN (RCNN) with active shape contours to segment the inner region of the LV; however, segmenting the outer region is much more challenging especially in apical views. Dong *et al.* [45] proposed a combination of a deformable contour method and a DL-based model for LV segmentation in 3-D echocardiography. Smistad and Østvik [27] developed a UNet-based model for 2-D echocardiograhpy images for LV segmentation, and Oktay *et al.* [46] extended this model for further improving the UNet-based segmentation. These models benefit from the knowledge of the expert and use prior knowledge for location-based guidance, for segmentation of the desired structure. These methods are helpful in cases where the DL model is not able to distinguish between two distinct objects having the same contextual and intensity information [47]. However, these methods require prior knowledge or morphological features and show poor real-time performance with high computing power. Jafari *et al.* [48] proposed a pipeline for real-time segmentation of the LV on a mobile device, taking the input from point of care devices and performing segmentation using a lightweight model based on the UNet.

1) Lightweight Semantic Segmentation Architectures: In general, segmentation models make use of an encoder–decoder structure, where the encoder is used to learn features and downsamples the input, whereas the decoder upsamples the output of the convolutions [31].

B. Minimizing Computational Complexity

Convolutions are performed by means of applying a filter to an input. However, different methods for executing such a filter can change the computational complexity of this operation. Two ways in which these filters can be adapted to decrease the number of parameters are as follows.

1) Filter Kernel Size: To reduce the amount of parameters, an effective method is to make use of filter stacking. Typically, this can be done by decomposing one $k \times k$ convolution layer into multiple 3×3 convolution layers. Previous works have shown that stacking three 3×3 convolution layers can have the same receptive field as a 7×7 convolution but with a decrease in parameters of 44.9% [49].

2) Filter Kernel Factorization: Another highly effective method is to replace $k \times k$ convolution kernels by multiple 1-D convolution kernels, such that they are decomposed in $1 \times k$ and $k \times 1$ kernels, which reduces computational cost effectively while maintaining the same receptive field as standard convolutions [43], [50]. The approach is very common and can be utilized to reduce the for any convolutional filter for reducing the computational complexity [51].

C. Convolution Designs

This section introduces several convolution designs widely used in lightweight models to reduce the number of parameters as compared with conventional networks.

1) Pointwise Convolution: This type of convolution consists of 1×1 projections and can be used to reduce the number of feature maps from the input [31]. These types of convolutions are often used in combination with 3×3 kernels to reduce the amount of output filters, leading to a decrease in computational complexity and, thus, the number of parameters [36].

2) Depthwise Separable Convolution: Typically, this type of convolution is characterized by a combination of convolutions over the input channel followed by a pointwise convolution [34]. The first step performs convolutions of size $d \times d$ over the input of size $i \times i$, which is performed over the number of input filters (C). At the end, a pointwise convolution is performed over the amount of input filters for the amount of output filters required (O). This reduces computational cost compared with regular convolutions by a factor (f) given by [36], [49]

$$f = 1/O + 1/d^2. \tag{1}$$

3) Dilated Convolutions: Several architectures, such as MiniNetV2 and FCdDN, make use of dilated convolutions in convolution blocks to learn multiscale features [33], [34], [43]. The dilation will enable the model to explore a wider feature space with fewer parameters needed. For example, a 3×3 kernel with a dilation rate (r) of 3 will resemble the feature space of a 7×7 kernel. The advantage of this approach is that more of the spatial information is retained compared with downsampling. Hence, this method can be used to minimize the amount of downsampling needed [43]. Dilation can also be used to create a multidilation depthwise separable convolution (MDSC) layer [34]. By performing this type of convolution operation, the number of parameters is reduced due to its larger receptive field in the feature space. In this case, the input is split over two channels. One channel performs a convolution over the input filters without dilation, and one channel performs convolution with dilation. These convolutions are added together after which a pointwise convolution is performed. An example is shown in Fig. 1.

D. State-of-the-Art Models

1) UNet: The UNet consists of a contracting path and an expanding path at different downsampled levels [52]. The contracting path consists of repeated pairs of 3×3 convolutions, followed by a rectified linear unit (ReLU) and a 2×2 max pool operation with stride 2 for downsampling. At each downsampling step, the number of feature channels is doubled. Every step in the expansive path consists of an upsampling of the feature map followed by a 2×2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. At the final layer, a 1×1 convolution is used to map each 64-component feature vector to the desired number of classes. The UNet architecture with its various blocks is shown in Fig. 2(a).

2) MiniNetV2: MiniNetV2 consists of different building blocks with several convolution modules incorporated [53]. The model is build as follows.

1) *Downsample Block:* Multiple 3×3 convolutional modules, downsample operations, and depthwise separable convolutions.
2) *Feature Extraction Block:* This is the most important module in the encoder. It provides multiple MDSCs with different dilation rates.
3) *Refinement Block:* Extracts additional features from the input image, which can assist in the refinement of previously learned features during the feature extractor block. This block is added after the feature extraction block.
4) *Upsample Block:* Upsamples the output with stride 2.

Compared with the original version of MiniNet [32], with 3.1 M parameters, MiniNetV2 consists of 0.52 M parameters and provides similar results. The model takes up a storage capacity of only 7.3 MB [34]. The MiniNetV2 architecture with its various blocks is shown in Fig. 2(b).

3) FCdDN: The network consists of 15 layers and is built up from different blocks [43]. The model is built as follows.

1) *1-D Dilated Layers:* Combines factorized convolution with a dilation rate of 2.
2) *Transition Down Blocks:* Downsamples the input by use of max pooling.
3) *Dense Dilated Blocks:* In these blocks, dense connectivity, dilated convolutions, and factorization are combined. This block consists of three dilation layers, with different dilation rates ($r = 2, 4, 8$) and a dropout of 0.2.

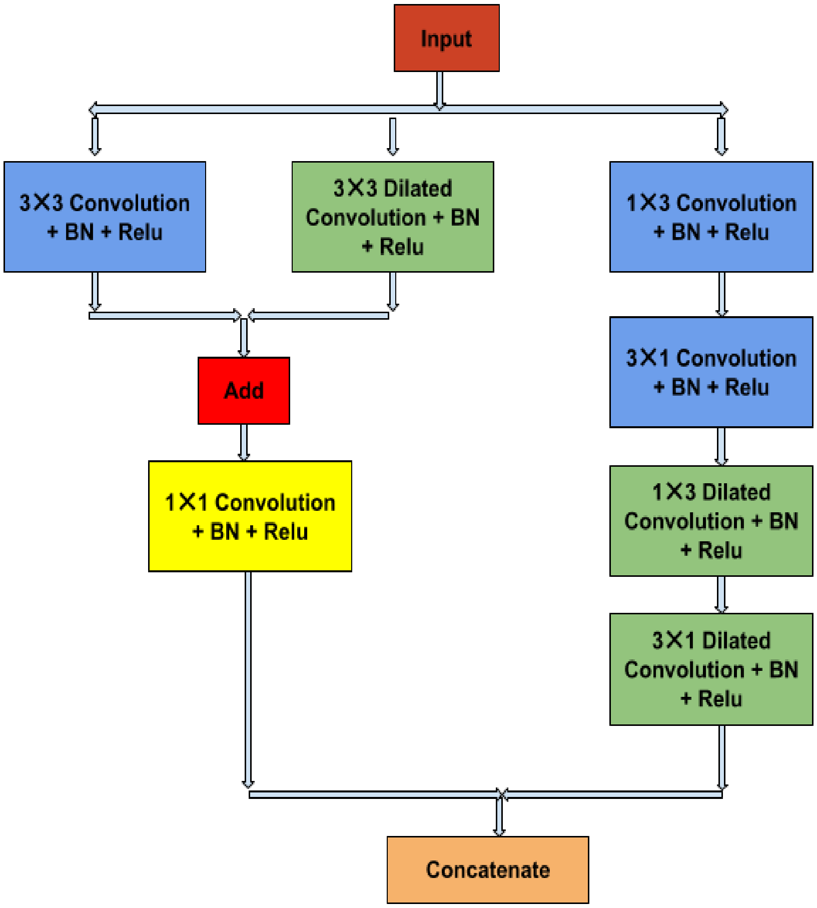

Input
3X3 Convolution + BN + Relu
3X3 Dilated Convolution + BN + Relu
1X3 Convolution + BN + Relu
Add
3X1 Convolution + BN + Relu
1X1 Convolution + BN + Relu
1X3 Dilated Convolution + BN + Relu
3X1 Dilated Convolution + BN + Relu
Concatenate

Fig. 1. Proposed split block with multidilation, depthwise, separable convolutions and factorized convolutions with the various kernel sizes and type of layer.

After every convolution, the output is concatenated with the previous input layer.

4) *Deconvolution Layers:* Used to upsample the output with strides of 2.

FCdDN is a lightweight segmentation model consisting of 0.2 M parameters and a model size of 3.4 MB, making it highly suitable for implementation in small devices. The FCdDN architecture with its various blocks is shown in Fig. 2(c).

III. Methods

A. LVNet (Proposed)

A new architecture is proposed, which is inspired by the FCdDN network, making use of both the dilation layers and the dense block. The proposed architecture is called LVNet. A new block (split block) (Fig. 1) is proposed, and the model architecture is visualized in Fig. 3. The description of the proposed split block and the model architecture is given in more detail in Sections III-A1 and III-A2.

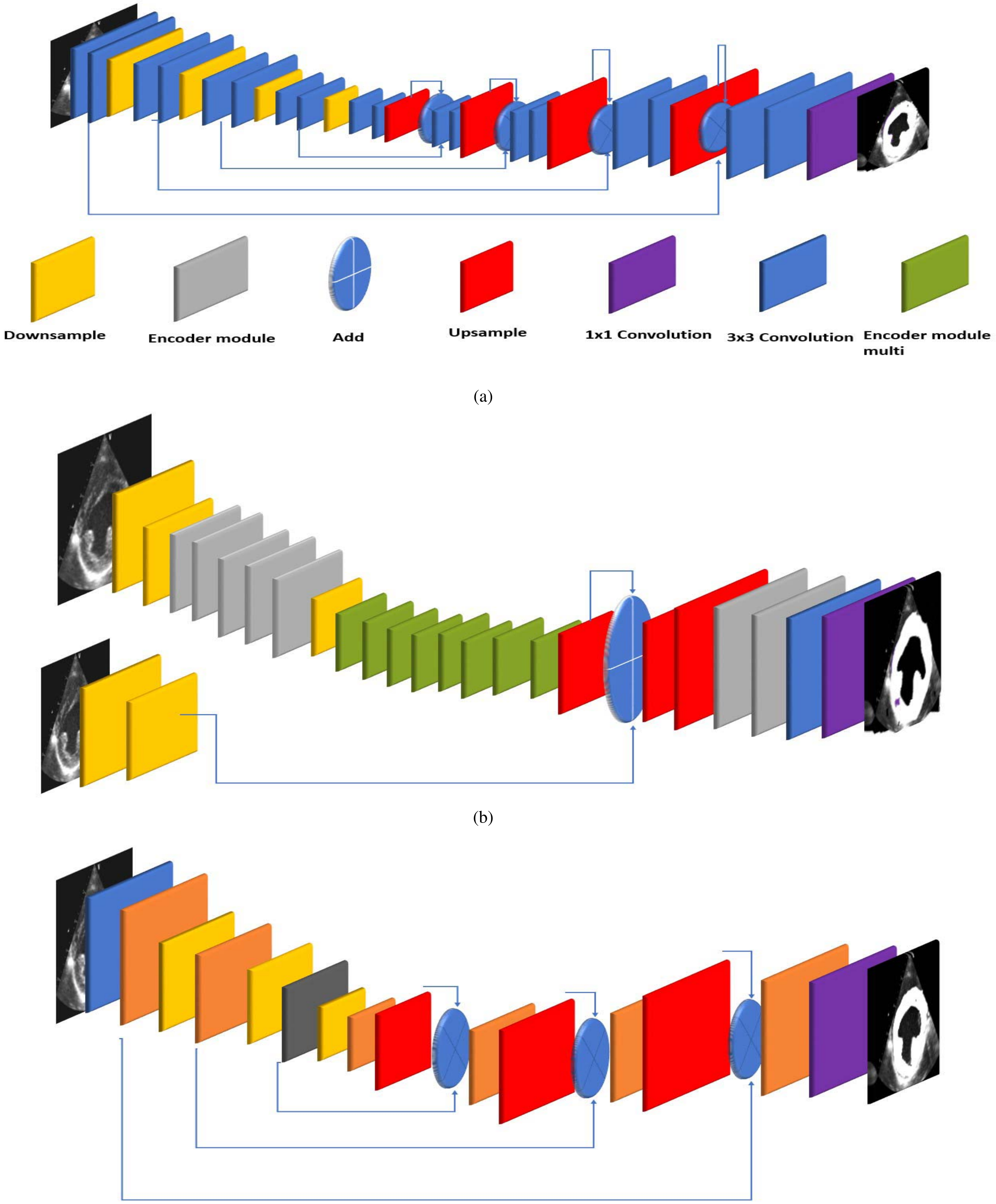
1) LVNet Architecture: LVNet first uses a 3 × 3 convolution over the input image. After this, the channels are split and fed into two split blocks, and the features are extracted in parallel. The motivation behind this is that it will help in extracting slightly different features in both channels, which enables the model to benefit from both channels' learning processes.

2) Split Block: The newly proposed block is called split block that is inspired by lightweight encoder–decoder network (LEDNet) [54]. In this split block, first two channels are created that acquire different features from the image.

1) One channel consists of two consecutive factorized convolutions. In the second factorized convolution, the dilation rate can be adapted, which helps in increasing the field of view without affecting the number of parameters.
2) The other channel is again factorized into two parallel convolutional channels—convolutional and dilated convolutional channels. The output of these channels is added, and finally, a pointwise convolution is performed.

The output of these two channels is concatenated and forms the output of this split block. This proposed block is different from the block utilized in LEDNet, as one of the channels consists of an MDSC instead of a factorized convolution to keep the number of parameters low while learning both local and global information [34]. At the end of the block, the channels are concatenated to combine features from both channels. The proposed block is portrayed in Fig. 1.

The channels consist of identical structures and consist of a split block differing in dilation rates. The split block

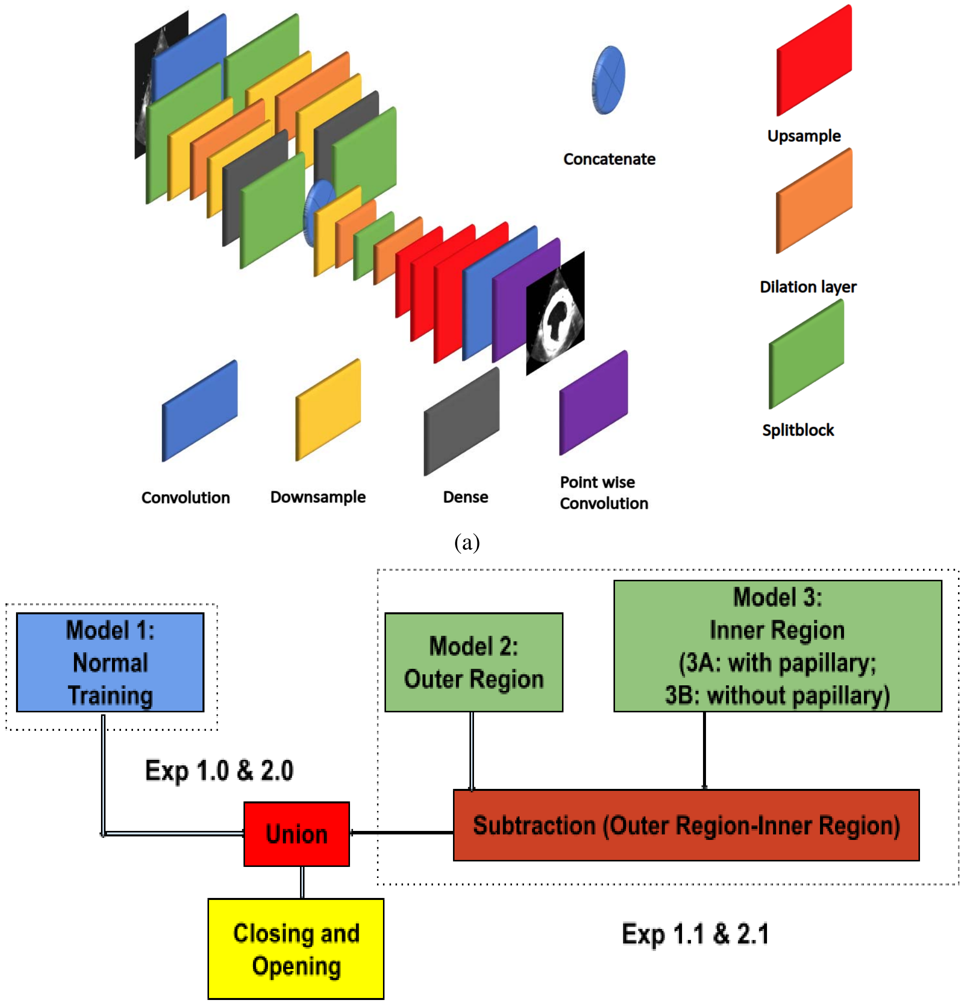


(a)

(b)

(c)

Fig. 2. Visualization of the UNet, MiniNetV2, and the FCdDN architectures and its different blocks. (a) Visualization of the UNet architecture and its different blocks. (b) Visualization of the MiniNetV2 architecture. The different blocks are shown in (a). (c) Visualization of the FCDDN architecture and its different blocks. The different blocks are shown in (a).



(a)

(b)

Fig. 3. Details of the proposed LVNet architecture. (a) Visualization of the proposed LVNet architecture (model architecture is shown in Table I) and its different blocks. (b) Post-processing pipeline utilized in this study combining Experiments 1.0 and 2.0.

uses the dilation rates of 2 and 3, respectively, in each branch. The small dilation rate of 2 in the first split block in the early layers will enable the model to efficiently extract features without losing too much of the local information, whereas large dilation rates will deprive too much local information [55]. This layer is followed by downsampling to reduce the size of the features. Next, a dilation layer is utilized, which consists of a dilation rate of 1, followed by another downsampling layer. It is followed by a dense block layer having the dilation rates of 2, 4, and 8, respectively. Finally, the channels are concatenated to utilize and fuse all the information from both the channels [56]. The middle layers of the model architecture consist of a dilation layer (dilation rate = 2), split block (dilation rate = 6), and another dilation layer (dilation rate = 2). Compared with FCdDN, dropout is not used, because this might lead to loss of important features, which could lead to a disadvantage in a model with a low amount of parameters. Finally, the result is upsampled by three consecutive transposed convolutions. At the end, first, a convolution and then a pointwise convolution are performed to place the output onto one feature map. The structure of LVNet is shown in Fig. 3(a), and the complete architecture of all the layers is given in Table I.

B. Experiments

The following different experiments were conducted to compare the performance of all models.

1) *Regular Training*: UNet, MiniNetV2, FCdDN, and the architecture proposed were all trained on the data for segmentation of the LV region WP and WOP muscles, which are referred to as Exp1.0 and Exp1.1, respectively. This is the strategy generally used for training of the segmentation models.
2) *Mask Subtraction:* A new strategy was proposed to further improve the segmentation outputs from the model. To improve the training process, the models were trained for the inner region WP muscle (endocardium border), the inner region WOP muscle, and for the outer region separately. It is expected that in this way, the model is able to learn lower level features, such as edges, more

TABLE I

LVNet Architecture Overview With the Type of the Layer, Input to Each Layer, the Number of Filters, and the Output Shape Corresponding to Each Layer

Layer	Type	Input	N_filters	Output shape
Input	-	-	-	256×256×1
l1	3×3 convolution	Input image	16	256×256×16
ls2	Splitblock (r=2)	l1	16	256×256×32
ls3	Downsample	ls2	16	128×128×16
ls4	Dilation layer	ls3	16	128×128×32
ls5	Downsample	ls4	32	64×64×32
ls6	Denseblock	ls5	32	64×64×32
ls7	Splitblock (r=3)	ls6	32	64×64×64
ls2.1	Splitblock (r=2)	l1	16	256×256×32
ls3.1	Downsample	ls2.1	16	128×128×16
ls4.1	Dilation layer	ls3.1	16	128×128×32
ls5.1	Downsample	ls4.1	32	64×64×32
ls6.1	Denseblock	ls5.1	32	64×64×32
ls7.1	Splitblock (r=3)	ls6.1	32	64×64×64
ls8	Concatenation	$ls7.1 \otimes ls7$	-	64×64×128
ls9	Downsample	ls8	64	32×32×64
ls10	Dilation layer (r=2)	ls9	64	32×32×128
ls12	Splitblock (r=6)	ls 11	64	32×32×128
ls13	Dilation layer (r=2)	ls12	64	32×32×192
ls14	Upsample	ls13	32	64×64×32
ls15	Upsample	ls14	16	128×128×16
ls16	Upsample	ls15	8	256×256×8
l17	3×3 convolution	ls16	64	256×256×64
output	Point-wise convolution	ls17	1	256×256×1

accurately, because it will be able to focus more on a specific area within the images. After training, the predicted inner mask was subtracted from the predicted outer mask. Simple processing was performed for all models by mapping any negative intensity value outside the final region after subtraction to 0, as this area is not relevant for the desired segmentation. These results can still contain oversegmented areas if the outer region segmentation results are poor. The result WP muscle is referred to as Exp2.0, while the result WOP muscle is referred to as Exp2.1.

C. LVNet + Post-Processing

Various post-processing pipelines have been proposed depending on the region to be segmented [57], [58]. The various techniques for post-processing include different types of morphological operations, contour refinement, and conditional random field-based post-processing methods [59]–[61]. Looney *et al.* [62] proposed a post-processing technique to remove the disconnected parts from the segmented region. They used binary dilation and erosion using a 3-D kernel having a radius of three voxels followed by a hole filling filter. This method removed the smaller regions and also filled the holes in the segmented area. Benjdira *et al.* [63] proposed a post-processing step using morphological operations for improving the segmentation quality of the images. It consists of four substeps for improving the segmentation consisting of removing small objects, removing the holes, and morphological closing followed by morphological opening. We also developed a post-processing pipeline based on morphological operations for further improving the segmentations obtained from the DL-based technique.

Exp 1.0 and Exp 2.0 denote the results of the normal training and training using mask subtraction WP muscles, while Exp 1.1 and Exp 2.1 denote the results of the normal training and training using mask subtraction WOP muscles. A new post-processing pipeline for LVNet is proposed, which takes the union of the results from regular training (Exp1.0 and Exp2.0) and the results from separate training after mask subtraction (Exp1.1 and Exp2.1). The post-processing pipeline is shown in Fig. 3(b). This creates an ensemble of models in which the models can enforce each other. After taking the union, morphological operations are performed, with a circular kernel of 7 × 7, consisting of closing and opening, consecutively. This fills small unsegmented areas in the outer region areas and removes small incorrectly segmented areas outside the outer region or inside the LV.

IV. Dataset and Metrics

A. Data Processing

An existing and fully annotated dataset of cardiac US image of the canine LV was reused in this study for training the models. These 2-D B-mode US (434 × 636 pixels) images originated from several US sequences in adult mongrel canines ($n = 13$) as reported in these previous studies [64], [65]. Animal handling was performed according to the Dutch law on animal experimentation and the European directive for the protection of animals used for scientific purposes (Directive 2010/63/EU). The protocol was approved by the Animal Experimentation Committee of Maastricht University. The US data were acquired using a GE Vivid5 US machine (GE Vingmed Ultrasound, Norten, Norway) [64], [65]. A GE PA2-5 phased array transducer (3-MHz center frequency, 75° opening angle, 90 frames/s) was used to image the short axis of the LV at the level of the papillary muscle.

The dataset was split into a training, validation, and test set, such that data from each canine are only present in one of the three sets, thereby making a canine-level split. The sequence length for each sample acquired varies from one canine to another, and the number of sequences also varies between 1 and 3. In total, the training data were derived from 16 sequences, while the validation and test datasets were derived from five sequences each. The training, validation, and test datasets consisted of 1445, 475, and 342 images, respectively. The images were resized to a resolution of 256 × 256 pixels.

Annotation was performed using a custom tool developed in MATLAB (The Mathworks Inc., Natick, MA, USA) for a previous study [65]. The endocardial boundary of the LV, including papillary muscle, the endocardial boundary WOP muscle, and the epicardial boundary, was segmented by three researchers and validated by two cardiac experts. The ground truth labels are made separately by three raters, and we did not performed the segmentation for the same image from the three raters. We did perform the validation analysis indepen-

TABLE II
TOTAL PARAMETERS INCLUDING THE MODEL PARAMETERS (TRAINABLE AND NON-TRAINABLE PARAMETERS), MODEL SIZE, AND THE FLOPS ARE SHOWN FOR ALL THE MODELS

Model	Total parameters	Trainable parameters	Non-trainable parameters	Model size (MB)	Flops (GF)
UNet	7,765,409	7,762,465	2,944	93.5	27.3
MiniNetV2	519,514	504,730	14,784	7.31	1.56
FCdDN	257,193	255,849	1,344	3.4	3.3
LVNet (Proposed)	332,217	328,137	4,080	4.87	3.34

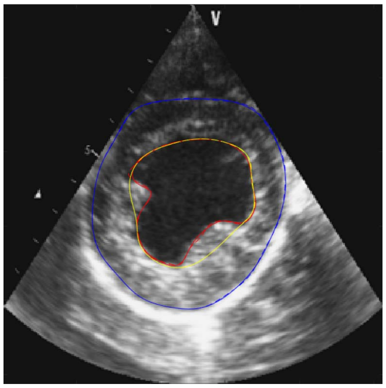


Fig. 4. Segmentation boundary of the outer region (shown in blue), the inner region WP muscles (shown in red), and the inner region WOP muscles (shown in yellow) for one example dataset.

dently by two experts, and if the experts do not agree, the segmentation was done again for those LV images. These boundary annotations are transformed to a complete mask for training the network. An example image is shown in Fig. 4.

B. Training Protocol

The training process was performed on a GPU (GeForce RTX 2080 Ti with 10.98-GB memory), with the same initial settings for all the models. The models were trained for a maximum of 300 epochs with batch size of 8, initial learning rate of 10^{-4}, and an Adam optimizer was utilized. To optimize the training, early stopping criteria on the validation set with a waiting period of 50 epochs, and learning rate reduction at a plateau for a waiting period of 20 epochs with a factor of 0.1, were utilized [66]. The loss function used for backpropagation is the Dice loss for all the models. The Dice loss considers the intersection between the prediction and the ground truth. Moreover, it is unaffected by the number of foreground or background pixels reducing the class-imbalance problem where learning algorithms reflect relatively low predictive accuracy concerning the infrequent class [67], [68].

C. Evaluation Measures

The main metric for evaluating the model performance is the DS, which is widely used for evaluating segmentation performance [1], [69]. The DS represents the overlap of the ground truth and the predicted segmentation output from the model

$$\mathrm{DS} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \tag{2}$$

Jaccard index (JI) or Intersection over Union (IOU) is another popular metric for measuring the degree of overlap between the two segmentation frames and is represented as [70]

$$\mathrm{JI} = \frac{\mathrm{DS}}{2 - \mathrm{DS}}. \tag{3}$$

Other metrics used are the sensitivity (SE) [71], specificity (SP) [71], and accuracy, serving as extra indicators of model performance [72]. The SE describes the portion of true positives (TPs) that are correctly classified by the model, with respect to the total number of positives [both TPs and false negatives (FNs)]

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{4}$$

The SP describes the portion of true negative (TN) values that are correctly classified with respect to the total number of negative values, i.e., the sum of the TNs and false positives (FP)

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{5}$$

Finally, accuracy describes the portion of correctly classified pixels from the total amount of classifications in an image

$$\mathrm{ac} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}. \tag{6}$$

V. RESULTS

A. Model Parameters

Table II describes the model parameters consisting of the total amount of parameters, trainable parameters, and non-trainable parameters with the model size for each of the models compared. The FCdDN network consists of the least amount of parameters, while UNet consists of the highest number of parameters and the model proposed lies in between. LVNet has a total of 1.6× less parameters (64% parameters) than MiniNetV2 resulting in a 1.5× lower memory usage (33.37% decrease in the model size). However, LVNet has 1.3× as many parameters as the FCdDN network at a 1.4× higher memory usage. The memory requirement of an LVNet model is only 5% as compared with a UNet model and only slightly higher than the FCdDN model. We also compared the floating point operations (Table II, column 6) required for running a single instance of the model. The proposed model requires the same number of flops compared with the FCdDN model while it requires more flops than the MiniNetV2 model.

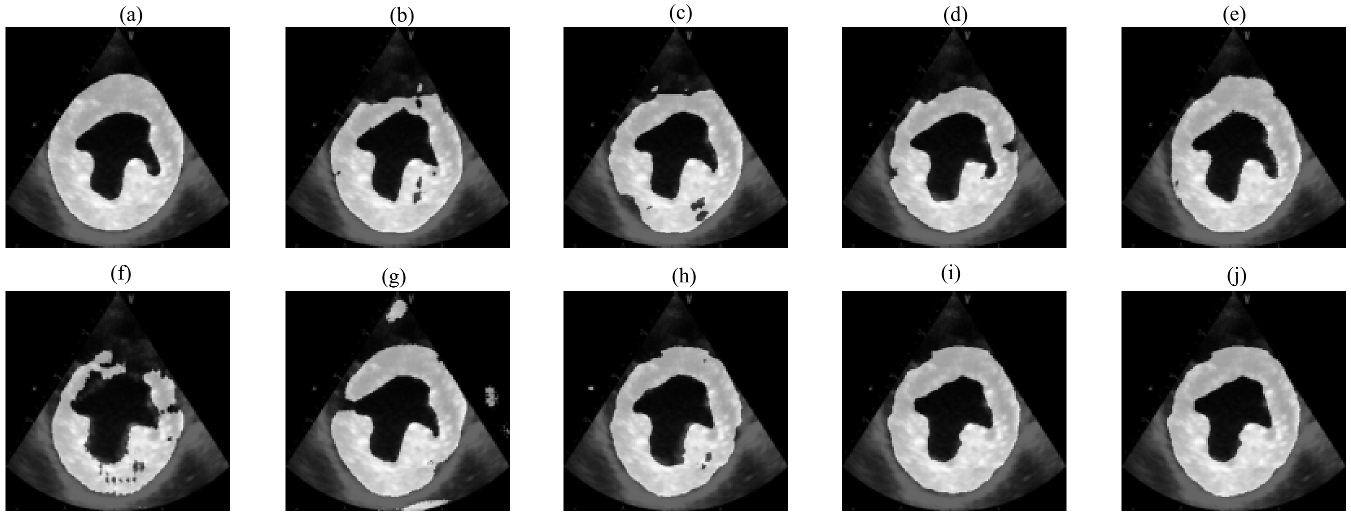


Fig. 5. Results of the segmentation WP muscle for different models. (a) Ground truth mask. (b) and (c) UNet. (d) and (e) MiniNetV2. (f) and (g) FCdDN. (h)–(j) LVNet.

TABLE III
PERFORMANCE OF THE DIFFERENT MODELS IN TERMS OF SE, SP, ACCURACY, DS, AND JI FOR THE SEGMENTATION WP MUSCLE FOR ALL THE MODELS FOR THE TEST DATASET

Model	Sensitivity	Specificity	Accuracy	Dice Score	Jaccard Index
UNet Exp1.0	0.928 ± 0.033	0.948 ± 0.025	0.945 ± 0.022	0.850 ± 0.046	0.743 ± 0.068
UNet Exp2.0	0.902 ± 0.070	0.902 ± 0.020	0.900 ± 0.020	0.768 ± 0.049	0.625 ± 0.064
MiniNetV2 Exp1.0	0.918 ± 0.036	0.951 ± 0.021	0.945 ± 0.016	0.851 ± 0.038	0.743 ± 0.057
MiniNetV2 Exp2.0	0.898 ± 0.070	0.920 ± 0.021	0.913 ± 0.017	0.805 ± 0.040	0.676 ± 0.057
FCdDN Exp1.0	**0.938 ± 0.043**	0.884 ± 0.035	0.890 ± 0.033	0.620 ± 0.105	0.457 ± 0.107
FCdDN Exp2.0	0.767 ± 0.058	0.898 ± 0.020	0.869 ± 0.026	0.717 ± 0.059	0.562 ± 0.073
LVNet Exp1.0	0.909 ± 0.057	**0.987 ± 0.011**	**0.982 ± 0.009**	0.851 ± 0.071	0.746 ± 0.101
LVNet Exp2.0	0.879 ± 0.051	0.978 ± 0.010	0.958 ± 0.013	0.894 ± 0.029	0.809 ± 0.046
LVNet Postprocessing	0.868 ± 0.052	0.985 ± 0.010	0.960 ± 0.013	**0.902 ± 0.028**	**0.823 ± 0.045**

B. Segmentation Including Papillary Muscle

The examples for the results of a segmentation including the papillary muscle are portrayed in Fig. 5 and the ground truth manual segmentation [Fig. 5(a)] performed by the experts. Fig. 5(b) and (c) visualizes the segmentation obtained using the UNet model for Exp1.0 and Exp2.0, respectively (for details of experiments, please refer Section III-B). As seen from the results, the segmentation results reveal gaps inside the mask as well as missing structure for the outer part. Fig. 5(d) and (e) visualizes the segmentation obtained using the MininetV2 model for Exp1.0 and Exp2.0, respectively. The results are better as compared with the UNet segmentation. Fig. 5(f) and (g) visualizes the segmentation obtained using the FcdDN model for Exp1.0 and Exp2.0, respectively. For the FCdDN model, the segmentation for Exp1.0 is incomplete, and the model is unable to identify all edges and can barely segment the outer region correctly.

Concerning Exp2.0, the segmentation of the inner region is more accurate as compared with Exp1.0; however, structures outside the desired region are included into the segmentation of the outer area, and thus, the shape is distorted. Comparing MiniNetV2 and LVNet, it can be observed that for Exp1.0, the MiniNetV2 had difficulty detecting the outer region correctly. In Exp2.0, both MiniNetV2 and LVNet mainly improved on the outer region. The inner structure regarding MiniNetV2 is elongated as compared with the ground truth, whereas for LVNet, edges are less sharp on the right wing of the inner region. Nevertheless, the overall shape matches the ground truth. Regarding post-processing, it can be seen that distorted parts are removed in the outer region for LVNet, and the inner shape is enhanced. At the same time in Fig. 5, it can be seen that for Exp1.0 for all models, the masks are incomplete and distorted except for LVNet. For MiniNetV2 and LVNet, Exp2.0 shows improvements in outer area with similar inner areas.

Table III shows the results for segmentation of the LV including the papillary muscle. Regarding Exp1.0, the DS and JI of LVNet perform in line with MiniNetV2 and UNet, while the LVNet model gives superior results as compared with all other models for Exp2.0. LVNet in combination with the post-processing pipeline proves to be superior in terms of SP, accuracy, DS, and JI as compared with the other model architectures reaching the values of 0.985, 0.960, 0.902, and 0.823, respectively.

C. Segmentation Excluding Papillary Muscle

Fig. 6 shows the result of a case for the segmentation without considering the papillary muscle. From the

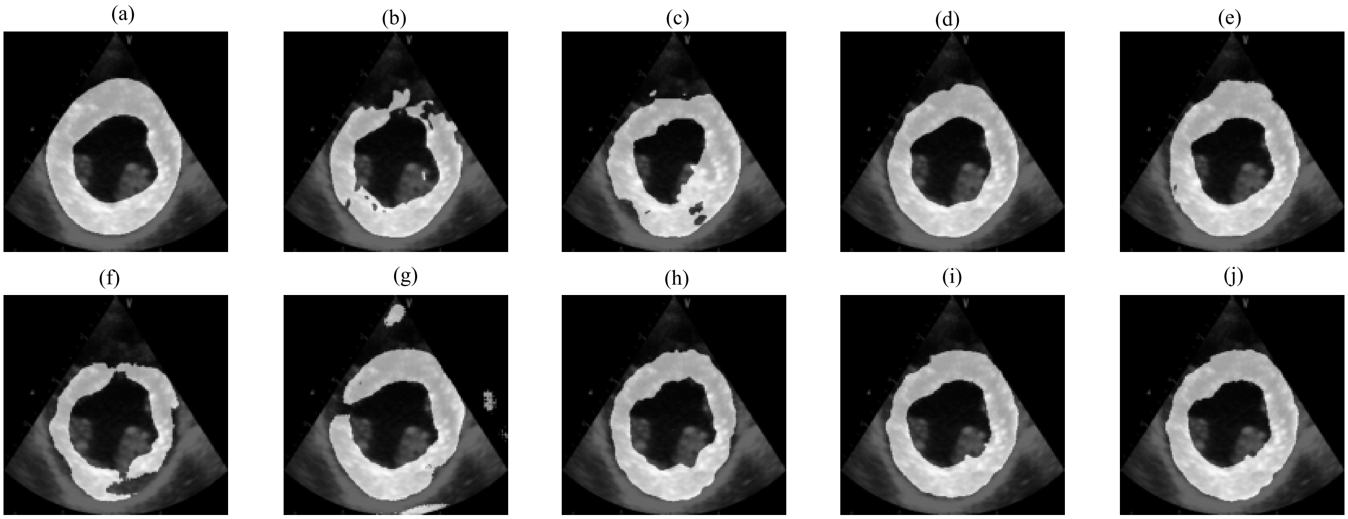


Fig. 6. Results of the segmentation WOP muscle for different models (a) Ground truth mask. (b) and (c) UNet. (d) and (e) MiniNetV2. (f) and (g) FCdDN. (h)–(j) LVNet.

TABLE IV
PERFORMANCE OF THE DIFFERENT MODELS IN TERMS OF SE, SP, ACCURACY, DS, AND JI FOR THE SEGMENTATION WOP MUSCLE FOR ALL THE MODELS FOR THE TEST DATASET

Model	Sensitivity	Specificity	Accuracy	Dice Score	Jaccard Index
UNet Exp1.1	**0.977 ± 0.018**	0.964 ± 0.015	0.965 ± 0.014	0.717 ± 0.207	0.560 ± 0.196
UNet Exp2.1	0.807 ± 0.088	0.915 ± 0.019	0.893 ± 0.026	0.739 ± 0.061	0.590 ± 0.074
MiniNetV2 Exp1.1	0.880 ± 0.053	0.970 ± 0.013	0.955 ± 0.016	0.870 ± 0.036	0.770 ± 0.053
MiniNetV2 Exp2.1	0.864 ± 0.075	0.926 ± 0.018	0.912 ± 0.018	0.785 ± 0.044	0.648 ± 0.060
FCdDN Exp1.1	0.902 ± 0.078	0.898 ± 0.043	0.890 ± 0.041	0.615 ± 0.146	0.458 ± 0.135
FCdDN Exp2.1	0.712 ± 0.078	0.894 ± 0.023	0.859 ± 0.029	0.654 ± 0.073	0.490 ± 0.083
LVNet Exp1.1	0.918 ± 0.082	**0.985 ± 0.008**	**0.981 ± 0.008**	0.865 ± 0.103	0.775 ± 0.134
LVNet Exp2.1	0.857 ± 0.057	0.978 ± 0.012	0.956 ± 0.014	**0.876 ± 0.033**	**0.781 ± 0.051**
LVNet Postprocessing	0.831 ± 0.060	0.983 ± 0.011	0.954 ± 0.013	0.874 ± 0.032	0.777 ± 0.049

visualization of the segmentation, it can be observed that LVNet and MiniNetV2 give similar visual appearance. Only for segmentation with FCdDN, it is noticeable that unwanted areas are segmented, and UNet yields undersegmented regions. Focusing on Fig. 6, it can be seen that for Exp1.1 for Unet and FCdDN, the masks are incomplete and distorted. In general, MiniNetV2 was able to capture the most complete outer area. Concerning Exp2.1, LVNet shows a more defined outer area, whereas the inner area is mostly closed. In this case, post-processing for LVNet is not considered beneficial.

Table IV shows the results for the segmentation of the LV excluding the papillary muscle. LVNet shows better results for SP, accuracy, DS, and JI than all other models for Exp2.1, while MininetV2 shows similar performance to LVNet. LVNet with the processing pipeline has proven to be powerful in terms of SP, accuracy, DS, and JI as compared with the other model architectures reaching the values of 0.985, 0.981, 0.865, and 0.775, respectively, and improving results with respect to Exp1.1.

VI. DISCUSSION

This study compares different lightweight segmentation models for their performance and proposes a new model and processing pipeline, which outperforms UNet, FCdDN, and MiniNetV2 for the trade-off between number of parameters and segmentation performance.

The results suggest that, in general, MiniNetV2 and LVNet perform similarly, both performing in the range of a Dice around 0.86–0.88 for segmentation WOP muscle, whereas FCdDN performs poorly in the range of 0.60–0.70 for most cases in segmentation WOP muscle. The total parameters and the model size for all the models are shown in Table II. In the case of segmentation WP muscle, for Exp1.0, LVNet outperforms MiniNetV2, while consisting of 187k parameters less than MiniNetV2. Besides, the model greatly improves upon FCdDN and UNet. The amount of parameters increased minimally as compared with FCdDN, which can be caused by the parallel feature extraction in the split channels and the split block but with an increased performance. Also, in the case of segmentation WOP muscle, for Exp1.1, MiniNetV2 performs similar to LVNet.

For Exp1.0, UNet and MiniNetV2 appear to perform better as compared with Exp2.0 (see Table III). Nevertheless, visually, Exp2.0 reflects a better outer region and an improved segmentation in most cases. However, the subtraction of masks during Exp2.0 will affect the outer region in cases of incorrectly segmented structures from the inner region. Hence, although the segmentation outputs are visually more attractive,

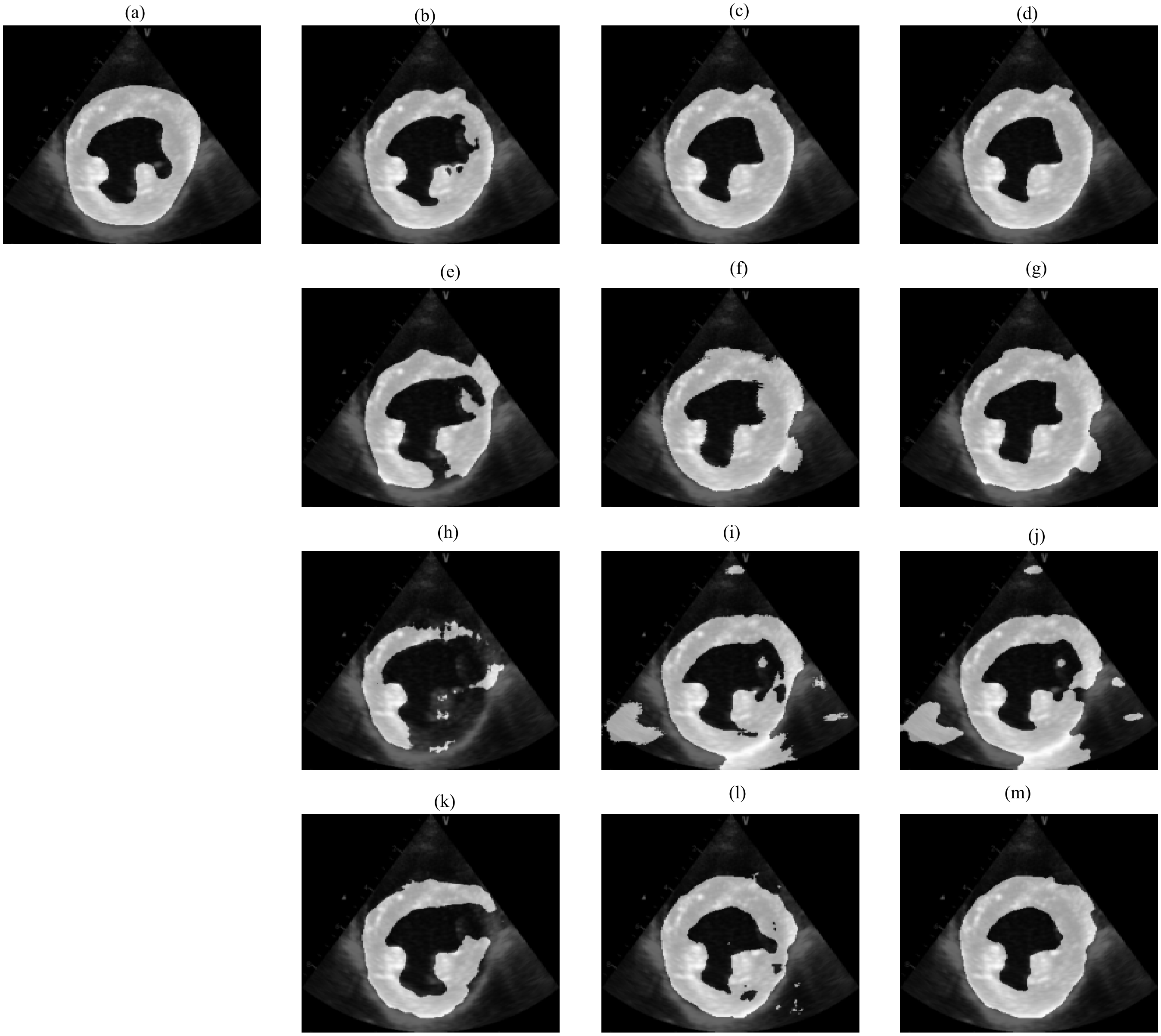


Fig. 7. Results of the segmentation WP muscle for different models showing the major effect of post-processing. (a) Ground truth mask. (b)–(d) UNet. (e)–(g) MiniNetV2. (h)–(j) FCdDN. (k)–(m) LVNet.

a lower DS might be reflected by an increase in FN values and a decrease in TP values [73]. Despite the fact that LVNet and MiniNetV2 yield high metrics, they both show their limitations in the segmentation. This happens mainly in cases where the inner area is small as compared with the outer area. This variability is revealed by the large standard deviations in the SE of all models. Meanwhile, FCdDN performs poorly as compared with the other models.

The UNet model performs better for Exps: 1.0 and 2.0 as compared with Exps: 1.1 and 2.1 (Tables III and IV). The reason behind this is the lack of a clear boundary separating the papillary muscles for the segmentation (Fig. 4). For cases with relatively small endocardial areas in the LV, segmentation of the inner region is poor for all models. As the target in this case is relatively small, as compared with the input size, downsampling can reduce performance of the model. The effect of downsampling is known to reduce spatial information; hence, the model might be unable to accurately capture the boundaries of the desired area in those cases [74]. This is in line with the findings that a good trade-off for the input resolution and training batch size reflects great impact on model performances [74].

A. Post-Processing

After performing post-processing on LVNet, DSs improved for segmentation WP muscle while decreased slightly for segmentation WOP muscle. The results reflect relatively low standard deviations, hence portraying more robust results than the other experiments, which is beneficial over the other models as was also observed in the literature [75]. Taking the union of Experiments 1.0 and 2.0 improved upon the results of UNet, LVNet, and MiniNetV2. For the segmentation WP muscles, a Dice of 0.90 was obtained, which shows excellent performance; meanwhile, a value of 0.87 was obtained for the segmentation WOP, which also reflects good predictive capacities. As both MiniNetV2 and LVNet performed similarly and LVNet performs significantly better than FCdDN, performing post-processing for LVNet will be more beneficial than for the other models, in terms of the trade-off between amount of parameters and performance. It is also more applicable to LVNet, because the number of parameters for using LVNet is 187k less (requiring only 64% parameters) than applying it for MiniNetV2, resulting in a much lower memory burden.

The results for another set of experiments for the segmentation WP muscle are shown in Fig. 7. Post-processing has mostly been proven to be successful in reducing the number of FN values [Fig. 7 (d), (g), (j), and (m)], by enhancing the outer region and structures around the edges of the inner area. However, what should be considered is that post-processing quality is limited by the least performing training method, mainly for the inner region. Thus, if for example the inner area is incompletely segmented by Exp2.0, a smaller inner area will be subtracted from the outer region, hence resulting in more FP values and thus reflecting a lower visual quality. Meanwhile, if the outer area is enlarged in one of the segmentation results, this will also lead to more FP values and, thus, decreases the SP of the approach. Similarly, if the outer area is undersegmented or inner area is oversegmented by Exp 2.0, a larger area will be subtracted from the outer region, hence resulting in more FN values and decreasing the SE of the segmentation technique. There are various other post-processing-based pipelines, which can be explored further for improving the results of the DL-based methods [59]–[61].

VII. Conclusion

For rapid diagnosis of abnormalities and to study LV function, the segmentation of LV is an important task. This study presented a lightweight CNN, LVNet, and a processing pipeline to segment the LV in US images. The method was demonstrated in a preexisting dataset. LVNet helps in rapid image analysis as it is a lightweight network, which can be readily deployed on low cost hardware and, hence, can be utilized in portable settings. Segmentation was performed for cases including and excluding the papillary muscle, outperforming MiniNetV2 for the former and showing equal performance for the latter. Moreover, LVNet greatly improved upon the FCdDN, which reflected a DS of 0.620. The model has proven to be more efficient in terms of amount of parameters and reduced memory usage more than MiniNetV2 and FCdDN. In addition, the post-processing pipeline has shown to improve the performance of LVNet, resulting in superior segmentation results, especially for less complex cases. In summary, as compared with UNet, FCdDN, and MiniNetV2, LVNet is considered more suitable for clinical implementation in terms of efficiency and performance, with the goal of accurate results and saving costs on memory usage. This also paves a way for lightweight models, which can be used for obtaining real-time results in low-cost settings.

References

[1] W.-Y. Hsu, “Automatic left ventricle recognition, segmentation and tracking in cardiac ultrasound image sequences,” *IEEE Access*, vol. 7, pp. 140524–140533, 2019.

[2] S. Dangi, Z. Yaniv, and C. A. Linte, “Left ventricle segmentation and quantification from cardiac cine MR images via multi-task learning,” in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Cham, Switzerland: Springer, 2018, pp. 21–31.

[3] Z. Zhuang, P. Jin, A. N. J. Raj, Y. Yuan, and S. Zhuang, “Automatic segmentation of left ventricle in echocardiography based on YOLOv3 model to achieve constraint and positioning,” *Comput. Math. Methods Med.*, vol. 2021, Mar. 2021, Art. no. 3772129.

[4] F. Khellaf, S. Leclerc, J. D. Voorneveld, R. S. Bandaru, J. G. Bosch, and O. Bernard, “Left ventricle segmentation in 3D ultrasound by combining structured random forests with active shape models,” *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105740J.

[5] M. Mignotte, J. Meunier, and J.-C. Tardif, “Endocardial boundary e timation and tracking in echocardiographic images using deformable template and Markov random fields,” *Pattern Anal. Appl.*, vol. 4, no. 4, pp. 256–271, Nov. 2001.

[6] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1987.

[7] J. A. Noble and D. Boukerroui, “Ultrasound image segmentation: A survey,” *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.

[8] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta, “Database-guided segmentation of anatomical structures with complex appearance,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 429–436.

[9] G. Carneiro and J. C. Nascimento, “Incremental on-line semi-supervised learning for segmenting the left ventricle of the heart from ultrasound data,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1700–1707.

[10] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, “Learning deep physiological models of affect,” *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.

[11] J. Vogel-Claussen *et al.*, “Left ventricular papillary muscle mass: Relationship to left ventricular mass and volumes by magnetic resonance imaging,” *J. Comput. Assist. Tomogr.*, vol. 30, no. 3, pp. 426–432, May 2006.

[12] M. Gao, C. Chen, S. Zhang, Z. Qian, D. Metaxas, and L. Axel, “Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles,” in *Proc. Int. Conf. Inf. Process. Med. Imag.* Berlin, Germany: Springer, 2013, pp. 184–195.

[13] K. Gilbert *et al.*, “Independent left ventricular morphometric atlases show consistent relationships with cardiovascular risk factors: A UK biobank study,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

[14] P. Rajiah, N. L. Fulton, and M. Bolen, “Magnetic resonance imaging of the papillary muscles of the left ventricle: Normal anatomy, variants, and abnormalities,” *Insights Imag.*, vol. 10, no. 1, pp. 1–17, 2019.

[15] S. Leclerc *et al.*, “Deep learning for segmentation using an open large-scale dataset in 2D echocardiography,” *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019.

[16] S. Leclerc, T. Grenier, F. Espinosa, and O. Bernard, “A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data,” in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[17] C. Chen *et al.*, “Deep learning for cardiac image segmentation: A review,” *Frontiers Cardiovascular Med.*, vol. 7, p. 25, Mar. 2020.

[18] S. Han *et al.*, “A deep learning framework for supporting the classification of breast lesions in ultrasound images,” *Phys. Med. Biol.*, vol. 62, no. 19, p. 7714, 2017.

[19] S. Liu *et al.*, “Deep learning in medical ultrasound analysis: A review,” *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019.

[20] F. Milletari *et al.*, “Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound,” *Comput. Vis. Image Understand.*, vol. 164, pp. 92–102, Nov. 2017.

[21] G. Carneiro, J. C. Nascimento, and 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.*, vol. 21, no. 3, pp. 968–982, Mar. 2012.

[22] K. G. Brown, D. Ghosh, and K. Hoyt, "Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 9, pp. 1820–1829, Sep. 2020.

[23] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Mischi, "Deep learning for super-resolution vascular ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1055–1059.

[24] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 325–336, Feb. 2018.

[25] B. Luijten *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020.

[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[27] E. Smistad, A. Ostvik, B. O. Haugen, and L. Lovstakken, "2D left ventricle segmentation using deep learning," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[28] L. Yu, Y. Guo, Y. Wang, J. Yu, and P. Chen, "Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1886–1895, Aug. 2017.

[29] G. Veni, M. Moradi, H. Bulu, G. Narayan, and T. Syeda-Mahmood, "Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 898–902.

[30] M. H. Jafari *et al.*, "A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Berlin, Germany: Springer, 2018, pp. 29–37.

[31] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*.

[32] J. Liu, Q. Li, R. Cao, W. Tang, and G. Qiu, "MiniNet: An extremely lightweight convolutional neural network for real-time unsupervised monocular depth estimation," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 255–267, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301544

[33] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNet: A light-weight context guided network for semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 1169–1179, 2021.

[34] I. Alonso, L. Riazuelo, and A. C. Murillo, "MiniNet: An efficient semantic segmentation ConvNet for real-time robotic applications," *IEEE Trans. Robot.*, vol. 36, no. 4, pp. 1340–1347, Aug. 2020.

[35] N. Awasthi, A. Dayal, L. R. Cenkeramaddi, and P. K. Yalavarthy, "Mini-COVIDNet: Efficient lightweight deep neural network for ultrasound based point-of-care detection of COVID-19," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 6, pp. 2023–2037, Jun. 2021.

[36] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[37] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[38] T. H. Mosbech, K. Pilgaard, A. Vaag, and R. Larsen, "Automatic segmentation of abdominal adipose tissue in MRI," in *Proc. Scand. Conf. Image Anal.* Ireland: Springer, 2011, pp. 501–511.

[39] J. Zhang, Y. Xie, P. Zhang, H. Chen, Y. Xia, and C. Shen, "Light-weight hybrid convolutional network for liver tumor segmentation," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 4271–4277.

[40] L. J. Ba and R. Caruana, "Do deep nets really need to be deep?" 2013, *arXiv:1312.6184*.

[41] H.-M. Zhang and B. Dong, "A review on deep learning in medical image reconstruction," *J. Oper. Res. Soc. China*, vol. 8, pp. 311–340, Jan. 2020.

[42] Y. He and T. Li, "A lightweight CNN model and its application in intelligent practical teaching evaluation," in *Proc. MATEC Web Conf.*, vol. 309. Les Ulis, France: EDP Sciences, 2020, p. 05016.

[43] A. Ouahabi and A. Taleb-Ahmed, "RETRACTED: Deep learning for real-time semantic segmentation: Application in ultrasound imaging," *Pattern Recognit. Lett.*, vol. 144, pp. 27–34, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865521000234

[44] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 17, 2021, doi: 10.1109/TPAMI.2021.3059968.

[45] S. Dong, G. Luo, K. Wang, S. Cao, Q. Li, and H. Zhang, "A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography," *BioMed Res. Int.*, vol. 2018, pp. 1–16, Sep. 2018.

[46] O. Oktay *et al.*, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2017.

[47] E. J. Rosana, C. Petitjean, P. Honeine, and F. Abdallah, "BB-UNet: U-Net with bounding box prior," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1189–1198, Oct. 2020.

[48] M. H. Jafari *et al.*, "Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 6, pp. 1027–1037, 2019.

[49] A. Briot, P. Viswanath, and S. Yogamani, "Analysis of efficient CNN design techniques for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 663–672.

[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[51] R. Rigamonti, A. Sironi, V. Lepetit, and P. Fua, "Learning separable filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2754–2761.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[53] I. Alonso, L. Riazuelo, and A. C. Murillo. (Apr. 2020). *Shathe/MiniNet-V2*. [Online]. Available: https://github.com/Shathe/MiniNet-v2

[54] Y. Wang *et al.*, "LEDNet: A lightweight encoder-decoder network for real-time semantic segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1860–1864.

[55] W. Jiang, Z. Xie, Y. Li, C. Liu, and H. Lu, "LRNNET: A light-weighted network with efficient reduced non-local operation for real-time semantic segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6.

[56] *The Difference Between the Add Layer and the Concatenate Layer in ShuffleNet—Programmer Sought*. Accessed: Jun. 15, 2021. [Online]. Available: https://www.programmersought.com/article/61336228767/

[57] P. Looney *et al.*, "Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 279–282.

[58] R. Almajalid, J. Shan, Y. Du, and M. Zhang, "Development of a deep-learning-based method for breast ultrasound image segmentation," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1103–1108.

[59] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, and B. Zhang, "Semantic segmentation of breast ultrasound image with fuzzy deep learning network and breast anatomy constraints," *Neurocomputing*, vol. 450, pp. 319–335, Aug. 2021.

[60] M. Martin, B. Sciolla, M. Sdika, X. Wang, P. Quetin, and P. Delachartre, "Automatic segmentation of the cerebral ventricle in neonates using deep learning with 3D reconstructed freehand ultrasound imaging," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[61] Y. Lei *et al.*, "Ultrasound prostate segmentation based on multi-directional deeply supervised V-Net," *Med. Phys.*, vol. 46, no. 7, pp. 3194–3206, Jul. 2019.

[62] P. Looney *et al.*, "Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning," *JCI Insight*, vol. 3, no. 11, pp. 1–9, Jun. 2018.

[63] B. Benjdira, K. Ouni, M. M. A. Rahhal, A. Albakr, A. Al-Habib, and E. Mahrous, "Spinal cord segmentation in ultrasound medical imagery," *Appl. Sci.*, vol. 10, no. 4, p. 1370, Feb. 2020.

[64] M. Strik *et al.*, "Interplay of electrical wavefronts as determinant of the response to cardiac resynchronization therapy in dyssynchronous canine hearts," *Circulat., Arrhythmia Electrophysiol.*, vol. 6, no. 5, pp. 924–931, Oct. 2013.

[65] L. S. Fixsen *et al.*, "Echocardiographic assessment of left bundle branch–related strain dyssynchrony: A comparison with tagged MRI," *Ultrasound Med. Biol.*, vol. 45, no. 8, pp. 2063–2074, Aug. 2019.

[66] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," *IEEE Access*, vol. 6, pp. 15844–15869, 2018.

[67] T. Cheng, X. Wang, L. Huang, and W. Liu, "Boundary-preserving mask R-CNN," in *Computer Vision—ECCV*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 660–676.

[68] C. X. Ling and V. S. Sheng, *Class Imbalance Problem*. Boston, MA, USA: Springer, 2010, p. 171, doi: 10.1007/978-0-387-30164-8_110.

[69] T. Eelbode *et al.*, "Optimization for medical image segmentation: Theory and practice when evaluating with dice score or Jaccard index," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3679–3690, Nov. 2020.

[70] P. Jaccard, "Distribution de la flore Alpine dans le Bassin des Dranses et dans quelques régions voisines," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 241–272, Jan. 1901.

[71] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity," *Brit. Med. J.*, vol. 308, p. 1552, Jun. 1994.

[72] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations," in *Proc. NESUG Health Care Life Sci.*, Baltimore, MD, USA, vol. 19, 2010, p. 67.

[73] A. Reinke *et al.*, "Common limitations of image processing metrics: A picture story," 2021, *arXiv:2104.05642*.

[74] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiol., Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190015.

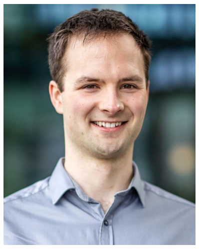
[75] M. B. Calisto and S. K. Lai-Yuen, "AdaEn-Net: An ensemble of adaptive 2D–3D fully convolutional networks for medical image segmentation," *Neural Netw.*, vol. 126, pp. 76–94, Jun. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608020300848

Navchetan Awasthi (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology (NIT) at Jalandhar, Jalandhar, India, in 2011, and the M.Tech. degree in computational science and the Ph.D. degree in medical imaging from the Indian Institute of Science (IISc), Bangalore, India, in 2016 and 2019, respectively.

He was a Research Fellow with the Massachusetts General Hospital, Boston, MA, USA, and Harvard University, Cambridge, MA, USA. He is currently a Postdoctoral Fellow with the Eindhoven University of Technology, Eindhoven, The Netherlands. His research interests include inverse problems in biomedical optics, medical image analysis, medical image reconstruction, biomedical signal processing, and deep learning.

Lars Vermeer received the International Baccalaureate degree in English and the Diploma degree from the 2College Durendael, Oisterwijk, The Netherlands, in 2018, and the bachelor's degree from the Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

His research interest include medical imaging techniques, medical imaging analysis, and deep learning.



Louis S. Fixsen received the M.Eng. degree in electrical and electronic engineering from the University of the West of England at Bristol, Bristol, U.K., in 2015, and the Ph.D. degree in biomedical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2021.

His research interests include ultrasound image analysis and heart failure.

Richard G. P. Lopata (Senior Member, IEEE) received the M.Sc. degree in biomedical engineering from the Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, in 2004, and the Ph.D. degree from Radboudumc, Nijmegen, The Netherlands, in 2010, with a focus on 2-D and 3-D ultrasound strain imaging: methods and in vivo applications.

He has been an Associate Professor with TU/e, heading the Photoacoustics and Ultrasound Laboratory Eindhoven (PULS/e lab) since 2014. The PULS/e lab facilitates research on technology development in the areas of ultrasound functional imaging, photoacoustics, and image-based modeling aimed to facilitate in and/or improve clinical decision-making for cardiovascular, musculoskeletal, and abdominal applications.

Josien P. W. Pluim (Fellow, IEEE) is currently a Professor of medical image analysis with the Eindhoven University of Technology, Eindhoven, The Netherlands, with a joint appointment at the University Medical Center at Utrecht, Utrecht, The Netherlands, for one day a week. She is also the Head of the Medical Image Analysis Group, Eindhoven University of Technology. Her research focus is on image analysis (e.g., registration, segmentation, detection, and machine/deep learning), both methodology development and clinical applications. The latter in particular targeted at neurology and oncology. Most of the research is performed in cooperation with clinical partners and/or industry.

Prof. Pluim was a member of the Executive Board of the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society. She is a fellow of the MICCAI Society. She was the Conference Chair of the Society of Photo-Optical Instrumentation Engineers (SPIE) Medical Imaging Image Processing from 2006 to 2009, the Chair of the Workshop on Biomedical Image Registration (WBIR) 2006, and the Program Co-Chair of MICCAI 2010. She serves/served as an Associate Editor for five journals, including the IEEE Transactions on Medical Imaging, the IEEE Transactions on Biomedical Engineering, *Medical Physics*, the *Journal of Medical Imaging*, and *Medical Image Analysis*.