Two-View Left Ventricular Segmentation and Ejection Fraction Estimation in 2D Echocardiograms

Frank Cally A. Tabuco[1]
fatabuco@up.edu.ph

Jose Donato A. Magno[2]
jamagno1@up.edu.ph

Nathaniel S. Orillaza, Jr.[3]
nsorillaza@up.edu.ph

Rani Ailyna V. Domingo[4]
rvdomingo@up.edu.ph

Prospero C. Naval, Jr.[1]
pcnaval@up.edu.ph

[1] Department of Computer Science,
University of the Philippines Diliman

[2] Division of Cardiovascular Medicine,
Department of Medicine,
UP College of Medicine -
Philippine General Hospital

[3] College of Medicine,
University of the Philippines Manila

[4] UP College of Medicine -
Philippine General Hospital

Abstract

Early detection of cardiovascular diseases through assessment of cardiac function is vital for accurate diagnosis, timely treatment, and improved prognosis. Among the various methods for estimation of ventricular function, 2D echocardiography remains to be one of the most valuable, accessible, and practical modalities in clinical practice. However, three main problems have persisted in the assessment of left ventricular (LV) ejection systolic function through ejection fraction (EF) measurement. First, current methods for analysis requires a series of procedures which are labor-intensive, time-consuming, and require high-level of skills to perform correctly. Second, semantic segmentation in 2D echocardiography often deals with low-quality, low-contrast images. Last, estimation of EF suffers from high *inter*-observer variability reaching as high as 14% error. To solve these problems, we developed segmentation and action recognition models in two-view 2D echocardiography for the automatic semantic segmentation of LV regions and estimation of LV EF. The segmentation model named channel-separated and dilated dense-Unet (CDDenseUnet) is capable of predicting segmented frames which outperformed current state-of-the-art architectures in terms of dice score, mean surface distance, and run-time performance reaching scores of 95.2%, 1.2mm, and 0.02 seconds, respectively. On the other hand, the prediction model named Two-Channel R(2+1)d is capable of analyzing segmented LV regions from echocardiogram videos in apical 2-chamber (A2C) and apical 4-chamber (A4C) views which produces better results than traditional estimation of EF reaching a mean absolute error of 3.8%. These new models have the potential to vastly improve LV EF measurement for the diagnosis of a wide variety of cardiac conditions and find great utility especially in complicated clinical scenarios or limited resource-settings where echocardiograms are prone to generation of sub-optimal image quality.

1 Introduction

Cardiovascular diseases (CVD) are serious illnesses which affect millions of people worldwide [12, 14]. CVD include several heart and blood vessel complications such as cerebrovascular disease, rheumatic heart disease, heart failure [57], and, even, invasive types like cardiac amyloidosis [41]. Physicians typically diagnose these diseases using medical imaging tools such as echocardiograms, magnetic resonance imaging scans, and computerized tomography scans [26, 41, 46, 48]. Among these imaging modalities, 2D echocardiography is commonly used to analyze the cardiac function of the LV regions [22, 32], since it is fast, cheap, and non-invasive [33, 41, 48]. With these modalities, along with growing research interests in this field, medical image datasets are now becoming increasingly common and publicly available with the purpose of improving analysis, detection, and treatment of various diseases.

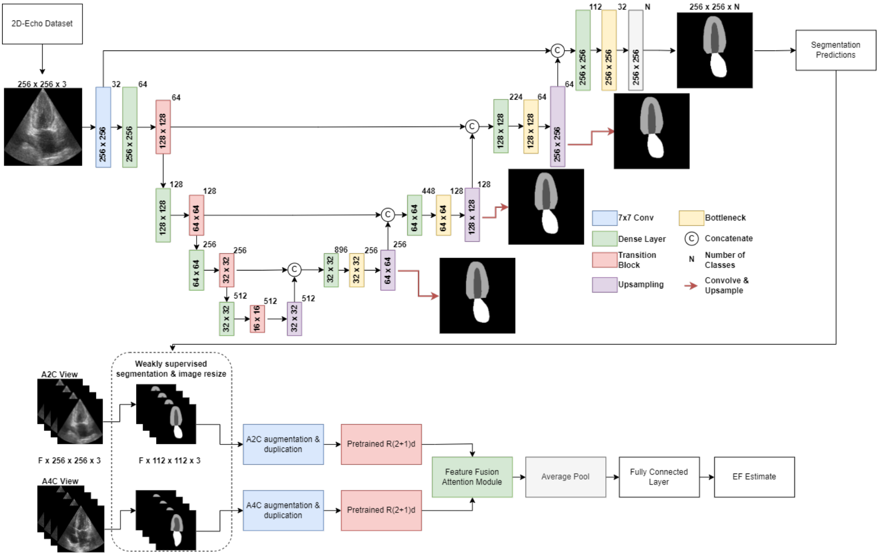

Figure 1: Proposed framework: CDDenseUnet (top) and TC-R(2+1)d (bottom).

Diagnosis of CVD typically requires a physician to analyze at least one cardiac cycle of an echocardiogram by tracing the borders of the LV endocardium before estimating EF [38, 57]. Not only is this labor intensive, but it is also time-consuming and requires high clinical skills to perform correctly [35]. Further, *inter*-observer variability is a major concern when measuring EF which could reach as high as 14% [24, 40, 47]. It is therefore critical to analyze echocardiography images and measure EF which minimizes variability, reduces segmentation errors, and processes automatic results in real-time.

In this research, segmentation and action recognition models were developed for the purpose of real-time, fully-automated, and accurate segmentation of LV regions and estimation of EF in A4C and A2C 2D echocardiography views (see Figure 1 for full framework). Our contributions in this work are:

- Developed a segmentation model which produces fast and accurate semantic segmentations of echocardiography images;

- Developed a two-channel action recognition model which uses fully-segmented LV regions in A2C and A4C cardiac views as inputs for the estimation of EF;

- Evaluated that when dealing with noisy, low-contrast, and sup-obtimal echocardiograms, it is necessary to estimate EF from fully-segmented frames;

- We believe that our overall framework is the first fully deep learning approach to achieve state-of-the-art performance in EF estimation superior to other works which used Modified Simpson's Rule.

The rest of the paper is structured as follows: section 2 discusses some related literature for semantic segmentation and action recognition models. The methods used for our models are discussed in detail in section 3. Section 4 gives an overview of the dataset used and experiment design. In section 5, we present the results of the experiments. Finally, the last sections discuss the research findings.

2 Review of Related Literature

2.1 Medical image segmentation

Segmentation is an important task for medical image analysis [18] especially for fields with difficult to perform imaging modalities, and those which produces low-quality images such as echocardiography. Despite the number of models created for medical image segmentation [16, 23, 42, 44, 54], LV segmentation in echocardiography still remains a prevailing problem. As mentioned in works by Leclerc *et al.* [29] and Liu *et al.* [31], echocardiogram videos are noisy and low-contrast resulting in difficult to differentiate tissue regions. Individual frames might also be incomplete with some edges, commonly the apex, mistakenly cropped-out during extraction by an echocardiographer resulting to sub-optimal images (see Figure 2).

The development of Unet [42] along with new techniques in improving label coherence such as squeeze & excitation modules [20], attention mechanisms [52], and atrous spatial pyramid pooling [7] paved way for the creation of several Unet variants which improve on its performance in specialized biomedical tasks. Related to our work, the success of attention mechanisms [52] in highlighting relevant spatial information served as the inspiration for the development of pyramid local attention network (PLANet) by Liu *et al.* [31]. Pyramid local attention and label coherence learning modules helped their model in learning context coherence for each pixels and its neighbors. With these mechanisms, PLANet achieved SOTA results for segmentation of the endocardium and epicardium outperforming architectures such as Unet [42], and Unet++ [56] as high as 96.2%, 1.5mm, and 4.6mm in terms of dice score, mean surface distance, and Hausdorff distance, respectively.

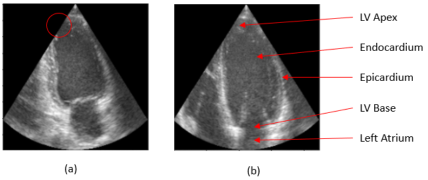

Figure 2: (a) Portion partially cropped-out during extraction in A2C cardiac view. (b) Parts of the left ventricle in A4C cardiac view. Image extracted from CAMUS dataset [29].

2.2 Ejection Fraction Estimation

The most common measurement for CVD diagnosis is to calculate LV EF or its pumping capacity within each cardiac cycle. This measurement has several diagnostic implications and is used by physicians to: 1) determine a patient's current status, 2) appropriately choose the correct treatment, and 3) check the treatment's effect on a patient [9, 32, 53]. Calculating EF requires quantification of endocardium LV volumes in end-systole (ES) and end-diastole (ED), and in at least two cardiac views, typically in A4C and A2C [27, 32, 43], using modified Simpson's rule [13, 57]. This is calculated as:

$$EF = 100 \times \frac{V_{ED} - V_{ES}}{V_{ED}} \tag{1}$$

Owing to the lack of large-scale datasets, the use of action recognition models to estimate EF was previously not possible. Therefore, EF estimation in deep learning traditionally relies on the performance of segmentation models in predicting masks of LV regions before Simpson's rule is used for EF measurement. Better segmentation models which produce more accurate segmentation masks provide better estimates of EF. The release of CAMUS [29] and EchoNet [37] datasets aims to solve this problem, and to increase the development of specialized segmentation and prediction models for echocardiography.

In the work by Ouyang *et al.* [38], they used a DeepLabv3 [7, 8] and R(2+1)d [50] for their EchoNet-Dynamic framework. The R(2+1)d model developed by Tran *et al.* [50] factorizes spatial and temporal kernels from a given video clip which helps the model learn relevant features within the spatial domain and across different timestamps [2, 38]. Results of their experiments showed that the beat-by-beat evaluation of EchoNet-Dynamic scored 4.05 mean absolute error (MAE), 5.32 root mean square error (RMSE), and 0.81 R^2. However, their work was only limited to A4C cardiac view which completely opposes the traditional approach, and only used segmentation for test-time augmentation. A dual-view network was also developed by Behnami *et al.* [5] which concatenates feature map results of a C3D [49] action recognition model to estimate EF. However, their approach completely skips the segmentation step and directly estimates EF from actual, low-quality frames. We will see in section 5 that the use of fully-segmented frames is crucial to produce more robust results, especially with sub-optimal images.

3 Methodology

3.1 Channel-separated and Dilated Dense-Unet

Dense convolution is the main characteristic and building block of our segmentation model. As mentioned in the work of Huang *et al.* [21], concatenation of feature maps allow dense convolution layers to retain and propagate important features throughout the network which are important for model learning. A standard DenseNet is composed of L layers, each with dense convolutions, batch normalization, pooling layers, and ReLU activation functions. While ResNet [17] introduced skip connections through addition of ResBlock outputs with its inputs, DenseNet further improves on these connections by concatenating layer outputs with preceding feature maps [55].

In this research, a dense block for the segmentation model is composed of three series of batch normalization, ReLU, and convolution blocks shown in Figure 3. Notable changes of our dense block against a standard dense block are the addition of dilation and group convolutions. Similar works by Cao *et al.* [6], Guan *et al.* [15], and Li *et al.* [30], added modified dense modules to Unet to tackle problems in tomography, microscopy, and tumor segmentation. The success of these models served as inspiration for the development of a specialized network for echocardiography.

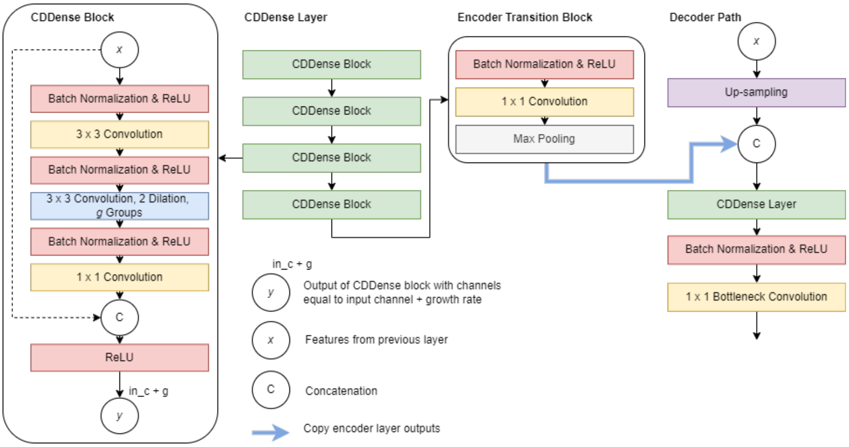
CDDense Block
x
Batch Normalization & ReLU
3 x 3 Convolution
Batch Normalization & ReLU
3 x 3 Convolution, 2 Dilation, g Groups
Batch Normalization & ReLU
1 x 1 Convolution
C
ReLU
in_c + g
y
CDDense Layer
CDDense Block
CDDense Block
CDDense Block
CDDense Block
Encoder Transition Block
Batch Normalization & ReLU
1 x 1 Convolution
Max Pooling
Decoder Path
x
Up-sampling
C
CDDense Layer
Batch Normalization & ReLU
1 x 1 Bottleneck Convolution
in_c + g
y
Output of CDDense block with channels equal to input channel + growth rate
x
Features from previous layer
C
Concatenation
Copy encoder layer outputs

Figure 3: Schematic and parts of CDDenseUnet.

The low-contrast, noisy, and variable nature of echocardiography images render standard convolution layers and the Unet architecture insufficient in learning small features necessary for the segmentation of LV regions [1]. With this, CDDenseUnet adopts two modifications in convolution layers to improve feature extraction, segmentation accuracy, and model inference time. The first modification implements group convolutions by filtering feature maps from an input layer using multiple kernels [25]. Group convolutions as described in the works of Howard *et al.* [19] and Tran *et al.* [51], reduces GPU compute requirements while increasing model accuracy. In this work, we add group convolutions by factoring the output channels with the middle channels, which is equivalent to a dense block's growth rate expressed as g. Similar to the work of Cao *et al.* [6], g is equal to the number of layer output channels divided by the layer size per dense layer, $\frac{L_{oc}}{L_s}$. Each dense layer in our model has a size of four dense blocks with output channels equal to 64, 128, 256, and 512. Another modification to our dense blocks is the addition of dilation to help extract semantic information among neighboring pixels. Each dense block has a dilation rate of d = 2 (see Figure 3 for implementation sample).

The combination of all these components comprises the proposed CDDenseUnet segmentation model shown in Figure 1. The model takes-in 256 x 256 resized echo images and initially passes through a 7 x 7 convolution layer. Succeeding feature extractions are in the form of dense layers (green), transition blocks (red), up-sampling layers (purple), and bottleneck blocks (yellow).

3.2 Two-Channel R(2+1)d

In this work, we extended the R(2+1) model to take-in two-view video clip inputs as seen in Figure 1. Prior to training the model, 256 x 256 resolution video clips were first segmented using the best model from CDDenseUnet. Our model takes-in input clips of shape 3 x N_F x 112 x 112, where N_F is a hyper-parameter that determines the number of frames to include from a video. Inputs for TC-R(2+1)d are 32-frame video clips with a *sampling stride = 2*. Due to the nature of the dataset used in this work, having less than 32-frames per video and only one cardiac cycle, we duplicated a video clip by stacking it nine times to imitate multiple cardiac cycles.

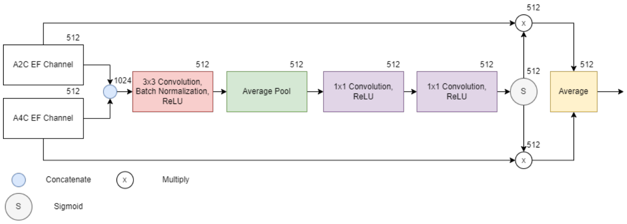

Figure 4: Feature Fusion Attention Module.

In order to estimate EF from two-views, we developed a feature fusion attention module (FFAM) which builds on the squeeze & excitation module by Hu *et al.* [20], and the feature fusion module by Liu *et al.* [31]. Channel-wise attention is extracted from concatenated feature maps of A2C and A4C branches. This weight is then multiplied to each branch feature maps before averaging the feature results. The full diagram is shown in Figure 4.

4 Experiment Setups

4.1 Dataset & Metrics

We evaluated our models using the Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) dataset released in 2019 by Lerlec *et al.* [29]. The dataset is composed of fully annotated echocardiography videos from 500 patients (50 of which are only accessible through the online platform) in two cardiac views for a total of 1000 videos. Ground truth masks of LV regions including the endocardium, epicardium, and left atrium in both ED and ES cardiac phases, and relevant cardiac assessments such as EF, ES volume, and ED volume were also provided to allow measurement of geometric and clinical performance for created models. Additionally, the dataset has different image qualities distributed into Good (175), Medium (230), and Poor (95). We made-use only of the 900 training videos provided, and followed the same training procedures of Leclerc *et al.* [29] where the dataset is split for 10-fold cross validation with 80% used for training, 10% for validation, and 10% for testing.

We used dice score, mean absolute distance (MAD), and 2D Hausdorff distance (HD) to evaluate the performance of CDDenseUnet. Similar to the work of Ouyang *et al.* [38], we used mean absolute error (MAE), root mean square error (RMSE), and R^2 to evaluate the EF estimates of TC-R(2+1)d. Additionally, we computed for correlation and bias for direct comparison to other EF estimation models.

For the above mentioned metrics, we used the MONAI [10] library to compute the segmentation geometric metrics, and scikit-learn [39] library to calculate the regression clinical metrics. We excluded poor quality frames during model testing since they are clinically useless as suggested by Leclerc *et al.* [29].

4.2 CDDenseUnet & TC-R(2+1)d setups

We used cross-entropy loss and dice loss to supervise training of the segmentation model (see equation 2). Deep supervision was also implemented during training which produces segmentation maps at different resolutions extracted from the decoder path. The final loss function is a weighted loss of all segmentation maps (see equation 3).

$$Loss_k = -\sum_{c=1}^{M} y_{gt,c} log(y_{pred,c}) + (1 - \frac{2 \times intersection}{y_{gt} + y_{pred}}) \tag{2}$$

$$Loss_F = Loss_1 + 0.5 \cdot (loss_2 + loss_3 + loss_4)/3 \tag{3}$$

We trained CDDenseUnet using an Adam optimizer with a learning rate of 10e^{-4}, and in batch sizes of 6. A cosine annealing scheduler is used to optimize the weights for all 60 epochs. During training, we applied data augmentation by random horizontal flip, elastic transform, random rotation from -10° to 10°, random scaling from 0.7 to 1.3, and random translation from -0.3 to 0.3.

Since EF estimation is framed as a regression problem, we used mean squared error loss to estimate the difference between the true value against the predicted estimate of EF. We trained TC-R(2+1)d using Adam optimizer with a learning rate of 10e^{-5}. We also implemented StepLR scheduler to decay the learning rate by 0.1 every 15 epochs. The model was trained for a total of 45 epochs at mini batch sizes of 5. Random translation was also used during training. All of our experiments were implemented using PyTorch and OpenCV using Google Colab Pro (Tesla P100).

5 Experiment Results & Discussion

Various segmentation algorithms were used in the experiments of Leclerc *et al.* [29] to tackle problems in the analysis of echocardiograms. These methods include non-deep learning algorithms such as structured random forests (SRF) [11, 28], and B-spline explicit active surface model (BEASM) [3, 4]. For deep learning algorithms, they trained different Unet [42] variants with changes in loss functions and architecture design, anatomically constrained neural networks (ACNN) [36], stacked hourglass (SHG) used for human pose estimation [34], and Unet++ [56]. Results of these methods are shown in Tables 1, 2, and 3. We also included other models for comparison such as PLANet [31], and residual-dilated Unet (ResDUnet) [1].

Performance results in Tables 1, 2, and 3 show that CDDenseUnet outperforms existing methods for 14 out of 19 endocardium, epicardium, and left atrium metric scores. Notable among these scores are the results for LV endocardium segmentation (the region typically used for EF estimation) where our model attained the highest score of 95.2% and 1.2mm for ED dice score and MAD, respectively. Additionally, inference time for segmentation prediction of a single frame is comparable in performace against lightweight models of Unet1 and Unet2 with an average time of 0.015 seconds on a Tesla P100 GPU. We tested the performance of Unet2 in the same computational environment and yielded an average runtime of 0.010 seconds for a single image. Comparing computation specifications with PLANet (runtime of 0.016 seconds) which used two Titan V GPUs, we evaluated our model on a Tesla

Table 1: Performance of CDDenseUnet against other segmentation methods, endocardium.

Model	ED			ES			Inference
	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	Time (s)
SRF[29]	0.895 ± 0.074	2.8 ± 3.6	11.2 ± 10.2	0.848 ± 0.137	3.6 ± 7.8	11.6 ± 13.6	-
BEASM-fully[29]	0.879 ± 0.065	3.3 ± 1.8	9.2 ± 4.9	0.826 ± 0.137	3.8 ± 2.1	9.9 ± 5.1	-
BEASM-semi[29]	0.920 ± 0.039	2.2 ± 1.2	6.0 ± 2.4	0.861 ± 0.070	3.1 ± 1.6	7.7 ± 3.2	-
Unet1[29]	0.934 ± 0.042	1.7 ± 1.0	5.5 ± 2.9	0.905 ± 0.063	1.8 ± 1.3	5.7 ± 3.7	0.090^a
Unet2[29]	0.939 ± 0.043	1.6 ± 1.3	5.3 ± 3.6	0.916 ± 0.061	1.6 ± 1.6	5.5 ± 3.8	0.140^a
ACNN[29]	0.932 ± 0.034	1.7 ± 0.9	5.8 ± 3.1	0.903 ± 0.059	1.9 ± 1.1	6.0 ± 3.9	-
SHG[29]	0.934 ± 0.034	1.7 ± 0.9	5.6 ± 2.8	0.906 ± 0.057	1.8 ± 1.1	5.8 ± 3.8	-
Unet++[29]	0.927 ± 0.046	1.8 ± 1.1	6.5 ± 3.9	0.904 ± 0.060	1.8 ± 1.0	6.3 ± 4.2	-
ResDUnet [1]	0.951 ± 0.030	1.4 ± 1.2	4.5 ± 1.2	-	-	-	-
PLANet [31]	0.951 ± 0.018	1.3 ± 0.5	**4.2 ± 1.4**	0.931 ± 0.032	1.4 ± 0.6	**4.3 ± 1.5**	0.016^b
CDDenseUnet (ours)	**0.952 ± 0.003**	**1.2 ± 0.1**	4.4 ± 0.3	**0.931 ± 0.003**	**1.2 ± 0.1**	4.4 ± 0.4	**0.015**^{c}

^{a}*Tesla M60,* ^{b}*Titan V,* ^{c}*Tesla P100*

P100 GPU which has weaker technical and performance specifications, thus inferring that our model is faster when measured in the same environment. Sample prediction contours of CDDenseUnet against ground truth contours are shown in Figure 5.

Table 2: Performance of CDDenseUnet against other segmentation methods, epicardium.

Model	ED			ES			Inference
	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	Time (s)
SRF[29]	0.914 ± 0.057	3.2 ± 2.0	13.0 ± 9.1	0.901 ± 0.078	3.5 ± 4.7	13.0 ± 11.1	-
BEASM-fully[29]	0.895 ± 0.051	3.9 ± 2.1	10.6 ± 5.1	0.880 ± 0.054	4.2 ± 2.0	11.2 ± 5.1	-
BEASM-semi[29]	0.917 ±0.038	3.2 ± 1.6	8.2 ± 3.0	0.900 ± 0.042	3.5 ± 1.7	9.2 ± 3.4	-
Unet1[29]	0.951 ± 0.024	1.9 ± 0.9	5.9 ± 3.4	0.943 ± 0.035	2.0 ± 1.2	6.1 ± 4.1	0.090
Unet2[29]	0.954 ± 0.023	1.7 ± 0.9	6.0 ± 3.4	0.945 ± 0.039	1.9 ± 1.2	6.1 ± 4.6	0.140
ACNN[29]	0.950 ± 0.026	1.9 ± 1.1	6.4 ± 1.4	0.942 ± 0.034	2.0 ± 1.2	6.3 ± 4.2	-
SHG[29]	0.951 ± 0.023	1.9 ± 1.0	5.7 ± 3.3	0.944 ± 0.034	2.0 ± 1.2	6.0 ± 4.3	-
Unet++[29]	0.945 ± 0.026	2.1 ± 1.0	7.2 ± 4.5	0.939 ± 0.034	2.1 ± 1.1	7.1 ± 5.1	-
PLANet [31]	0.962 ± 0.012	1.5 ± 0.5	**4.6 ± 1.5**	**0.956 ± 0.014**	1.6 ± 0.6	**4.6 ± 1.4**	0.016
CDDenseUnet (ours)	**0.962 ± 0.002**	**1.5 ± 0.0**	5.1 ± 0.3	0.954 ± 0.002	**1.6 ± 0.1**	5.5 ± 0.2	**0.015**

Table 3: Performance of CDDenseUnet against other segmentation methods, left atrium.

Model	ED			ES			Inference
	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	↑ Dice Score	↓ MAD (mm)	↓ HD (mm)	Time (s)
Unet1[29]	0.889	2.2	5.7	0.918	2.0	5.3	0.090
Unet2[29]	0.848	2.6	6.9	0.888	2.1	6.2	0.140
ACNN[29]	0.881	2.3	6.0	0.911	2.2	5.8	-
CDDenseUnet (ours)	**0.891**	**1.5**	**5.6**	**0.921**	**1.4**	**5.2**	**0.015**

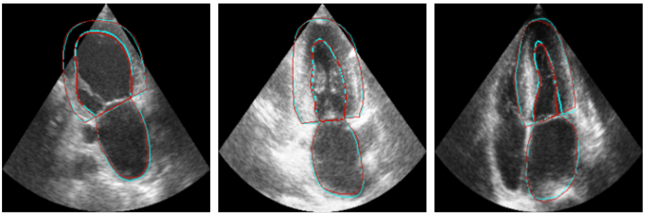

Figure 5: Prediction (blue) contours against ground truth (red). Right-most in A4C view.

For EF estimation, we grouped the different existing methods into five categories namely: 1) cardiologists estimates; 2) non-deep learning segmentation methods with traditional EF estimation; 3) deep learning segmentation with traditional EF estimation; 4) actual frames

Table 4: Performance of TC-R(2+1)d against other existing methods for EF estimation.

Technique	Observer/Model	↑ Correlation	bias ± σ	↓ MAE (%)	↓ RMSE (%)	↑ R^2
Cardiologists	O1a vs O2 (*inter*-observer)[29]	0.801	-9.1 ± 8.1	10.0	-	-
	O1a vs O3 (*inter*-observer)[29]	0.646	-12.6 ± 10.0	13.4	-	-
	O2 vs O3 (*inter*-observer)[29]	0.569	3.5 ± 11.0	8.5	-	-
	O1a vs O1b (*intra*-observer)[29]	0.896	-2.3 ± 5.7	**0.9**	-	-
Non-deep learning segmentation + Simpson's Rule	SRF[29]	0.465	-11.5 ± 15.4	12.8	-	-
	BEASM-fully[29]	0.731	-9.8 ± 8.3	10.7	-	-
	BEASM-semi[29]	0.790	-9.4 ± 7.2	10.0	-	-
Deep learning segmentation + Simpson's Rule	Unet1[29]	0.791	-0.5 ± 7.7	5.6	-	-
	Unet2[29]	0.823	-1.0 ± 7.1	5.3	-	-
	ACNN[29]	0.799	-0.8 ± 7.5	5.7	-	-
	SHG[29]	0.770	-1.4 ± 7.8	5.7	-	-
	Unet++[29]	0.789	-1.8 ± 7.7	5.6	-	-
	Automated EF [45]	-	1.8 ± 8.9	6.7	-	-
	PLANet [31]	0.882	0.6 ± 5.8	-	-	-
Actual frames + action recognition	R(2+1)d (A2C-only)	0.689	-0.3 ± 8.4	6.8	8.6	0.425
	R(2+1)d (A4C-only)^a	0.705	-0.3 ± 8.4	6.7	8.5	0.428
	TC-R(2+1)d (ours) - concatenate^b	0.777	-0.5 ± 7.4	5.9	7.6	0.535
	TC-R(2+1)d (ours)	0.786	-0.3 ± 7.3	5.8	7.5	0.550
Deep learning segmentation + action recognition	R(2+1)d (A2C-only)	0.827	0.1 ± 6.4	5.2	6.6	0.645
	R(2+1)d (A4C-only)	0.793	-0.5 ± 7.2	5.9	7.5	0.548
	TC-R(2+1)d (ours) - concatenate	0.879	0.1 ± 5.5	4.4	5.5	0.718
	TC-R(2+1)d (ours) - endocardium^c	0.897	-0.1 ± 5.0	4.0	5.1	0.790
	TC-R(2+1)d (ours)	**0.903**	-0.8 ± 4.9	**3.8**	**5.0**	**0.792**

[a] *Similar to EchoNet-Dynamic [38].*
[b] *Replaces FFAM with concatenation. Similar to Dual-View EF [5].*
[c] *Using endocardium masks only for EF estimation.*

with action recognition model; and 5) deep learning segmentation with action recognition model. Our work belongs to the fifth category where we used the best performing CDDense-Unet model to predict LV region segmentations for a given 32-frame video clip, and used TC-R(2+1)d to estimate EF. We compared the performance of our model against these methods in Table 4. Based on the suggestion of Leclerc *et al.* [29], bias was not considered as an evaluation metric since a lower bias score does not entail a better model. Of the relevant metrics, TC-R(2+1)d achieved the best values for correlation, RMSE, and R^2 with scores of 0.903, 5.0%, and 0.792, respectively. The model also greatly outperformed *inter*-observer estimates and previous models in MAE with a score of 3.8% and only behind *intra*-observer estimate error of 0.9%. Further, the model has an average run-time performance of 0.02 seconds for estimating EF in a single two-view video clip.

The results in Table 4 also show comparative results for estimating EF when using actual frames against segmented frames, and single-view versus two-view estimates. For experiments using single-view clips, only a R(2+1)d network is used without the FFAM module, while segmentation was skipped during pre-processing of clips for actual frames. We trained these ablation experiments using the same settings discussed in Section 4. The results show that the models produced higher scores when using segmented frames than actual video clips. This is due to more visible expansions and contractions of LV regions, which are difficult to visualize in actual, low-quality frames. Additionally, the use of two-view and fully-segmented LV regions, instead of endocardium masks only, notably increased model performance by as much as 0.2 MAE percentage points.

Based on these experiments, the use of action-recognition models alone will not yield great performance. Previous deep learning approaches such as EchoNet [38] and Dual-View EF [5] pale in comparison to our work and even with the traditional approch of using Simpson's Rule reaching MAE scores of only 6.7% and 5.9%, respectively. Therefore, we argue that segmentation of LV regions is an important step towards achieving lower error rates especially when dealing with noisy, low-quality, and sub-optimal echocardiograms. The use of two-view segmented video clip inputs in low-quality echocardiography produces better EF estimation results than actual and/or single-view videos, exceeding the scores of PLANet [31] and Unet2 [29].

6 Conclusion

In this research, we proposed a deep learning framework specifically designed to tackle challenges in echocardiography such as low-contrast images, time-consuming procedures, and high *inter*-observer variability. The framework which is composed of a deep segmentation model, CDDenseUnet, and an action recognition model, TC-R(2+1)d, showed improved performance against existing methods in relevant geometric and clinical metrics reaching the best scores of 95.2% dice, and 3.8% MAE, respectively, and with faster run-time performance. Although it is possible to estimate EF using actual frames and in single-view, the use of two-view and fully-segmented echocardiogram frames yielded better results superior to other existing methods.

Acknowledgements The main author is supported by a scholarship grant from the Department of Science and Technology - Engineering and Research Development for Technology (DOST-ERDT). This work is a collaboration with the University of the Philippines College of Medicine and University of the Philippines - Surgical Innovation and Biotechnology Laboratory (UP SIBOL).

References

[1] Alyaa Amer, Xujiong Ye, and Faraz Janan. Resdunet: A deep learning-based left ventricle segmentation method for echocardiography. *IEEE Access*, 9:159755–159763, 2021. doi: 10.1109/ACCESS.2021.3122256.

[2] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6816–6826, 2021. doi: 10.1109/ICCV48922.2021.00676.

[3] D. Barbosa, T. Dietenbeck, B. Heyde, H. Houle, D. Friboulet, J. D'hooge, and O. Bernard. Fast and fully automatic 3d echocardiographic segmentation using b-spline explicit active surfaces. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1088–1091, 2012. doi: 10.1109/ISBI.2012.6235748.

[4] Daniel Barbosa, Thomas Dietenbeck, Joel Schaerer, Jan D'hooge, Denis Friboulet, and Olivier Bernard. B-spline explicit active surfaces: An efficient framework for real-time 3-d region-based segmentation. *IEEE Transactions on Image Processing*, 21(1): 241–251, 2012. doi: 10.1109/TIP.2011.2161484.

[5] Delaram Behnami, Zhibin Liao, Hany Younan Azer Girgis, Christina Luong, Robert Rohling, Kenneth Gin, Teresa Tsang, and Purang Abolmaesumi. *Dual-View Joint Estimation of Left Ventricular Ejection Fraction with Uncertainty Modelling in Echocardiograms*, pages 696–704. 10 2019. ISBN 978-3-030-32244-1. doi: 10.1007/978-3-030-32245-8_77.

[6] Yue Cao, Shigang Liu, Yali Peng, and Jun Li. Denseunet: densely connected unet for electron microscopy image segmentation. *IET Image Processing*, 14(12):2682–2689, 2020. doi: https://doi.org/10.1049/iet-ipr.2019.1527. URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2019.1527.

[7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.

[8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. doi: 10.1109/TPAMI.2017.2699184.

[9] John G. F. Cleland and James McGowan. Heart failure due to ischaemic heart disease: epidemiology, pathophysiology and progression. *Journal of Cardiovascular Pharmacology*, 33, 1999.

[10] MONAI Consortium. Monai: Medical open network for ai. doi: 10.5281/zenodo.4323058. URL https://monai.io.

[11] Piotr Dollár and C. Lawrence Zitnick. Fast edge detection using structured forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1558–1570, 2015. doi: 10.1109/TPAMI.2014.2377715.

[12] Susan A. Everson-Rose and Tené T. Lewis. Psychosocial factors and cardiovascular diseases. *Annual Review of Public Health*, 26(1):469–500, 2005. doi: 10.1146/annurev.publhealth.26.021304.144542. URL https://doi.org/10.1146/annurev.publhealth.26.021304.144542. PMID: 15760298.

[13] ED Folland, AF Parisi, PF Moynihan, DR Jones, CL Feldman, and DE Tow. Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques. *Circulation*, 60(4), 1979.

[14] Thomas A. Gaziano, Asaf Bitton, Shuchi Anand, Shafika Abrahams-Gessel, and Adrianna Murphy. Growing epidemic of coronary heart disease in low- and middle-income countries. *Current problems in cardiology*, 35(2):72–115, 2010. doi: doi: 10.1016/j.cpcardiol.2009.10.002.

[15] Steven Guan, Amir A. Khan, Siddhartha Sikdar, and Parag V. Chitnis. Fully dense unet for 2-d sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health Informatics*, 24(2):568–576, Feb 2020. ISSN 2168-2208. doi: 10.1109/jbhi.2019.2912935. URL http://dx.doi.org/10.1109/JBHI.2019.2912935.

[16] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1748–1758, 2022. doi: 10.1109/WACV51458.2022.00181.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778, 06 2016. doi: 10.1109/CVPR.2016.90.

[18] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of Digital Imaging*, 32, 05 2019. doi: 10.1007/s10278-019-00227-x.

[19] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017. URL https://arxiv.org/abs/1704.04861.

[20] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011–2023, 2020. doi: 10.1109/TPAMI.2019.2913372.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. doi: 10.1109/CVPR.2017.243.

[22] Institute of Medicine (US) Committee on Social Security Cardiovascular Disability Criteria. *Ischemic Heart Disease*, chapter 7. Washington (DC), 2010. URL https://www.ncbi.nlm.nih.gov/books/NBK209964/.

[23] Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation, 2020. URL https://arxiv.org/abs/2006.04868.

[24] Ateet Kosaraju, Amandeep Goyal, Yulia Grigorova, and Amgad N. Makaryus. *Left Ventricular Ejection Fraction*. StatPearls Publishing, Treasure Island, FL, 2021. URL https://www.ncbi.nlm.nih.gov/books/NBK459131/.

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[26] Claes Ladefoged, Philip Hasbak, Charlotte Hornnes, Liselotte Højgaard, and Flemming Andersen. Low-dose pet image noise reduction using deep learning: application to cardiac viability fdg imaging in patients with ischemic heart disease. *Physics in Medicine & Biology*, 66, 02 2021. doi: 10.1088/1361-6560/abe225.

[27] Roberto M. Lang, Luigi P. Badano, Victor Mor-Avi, Jonathan Afilalo, Anderson Armstrong, Laura Ernande, Frank A. Flachskampf, Elyse Foster, Steven A. Goldstein, Tatiana Kuznetsova, Patrizio Lancellotti, Denisa Muraru, Michael H. Picard, Ernst R. Rietzschel, Lawrence Rudski, Kirk T. Spencer, Wendy Tsang, and Jens-Uwe Voigt. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the american society of echocardiography and the european association of cardiovascular imaging. *Journal of the American Society of Echocardiography*,

28(1):1–39.e14, 2015. ISSN 0894-7317. doi: https://doi.org/10.1016/j.echo.2014.10.003. URL https://www.sciencedirect.com/science/article/pii/S0894731714007457.

[28] Sarah Leclerc, Thomas Grenier, Florian Espinosa, and Olivier Bernard. A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2d echocardiographic data. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2017. doi: 10.1109/ULTSYM.2017.8092797.

[29] Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Ostvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Berg, Pierre-Marc Jodoin, T. Grenier, Carole Lartizien, Jan Drhooge, Lasse Løvstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, PP:1–1, 02 2019. doi: 10.1109/TMI.2019.2900516.

[30] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37(12):2663–2674, 2018. doi: 10.1109/TMI.2018.2845918.

[31] Fei Liu, Kun Wang, Dan Liu, Xin Yang, and Jie Tian. Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography. *Medical Image Analysis*, 67:101873, 01 2021. doi: 10.1016/j.media.2020.101873.

[32] Ahmad Malik, Daniel Brito, Sarosh Vaqar, and Lovely Chhabra. *Congestive Heart Failure*. StatPearls Publishing, Treasure Island, FL, 2021. URL https://www.ncbi.nlm.nih.gov/books/NBK430873/.

[33] Dania Mohty, Thibaud Damy, Pierre Cosnay, Najmeddine Echahidi, Danielle Casset-Senon, Patrice Virot, and Arnaud Jaccard. Cardiac amyloidosis: Updates in diagnosis and management. *Archives of Cardiovascular Diseases*, 106(10):528–540, 2013. ISSN 1875-2136. doi: https://doi.org/10.1016/j.acvd.2013.06.051. URL https://www.sciencedirect.com/science/article/pii/S187521361300274X.

[34] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation, 2016. URL https://arxiv.org/abs/1603.06937.

[35] Walter G. O'Dell. Accuracy of left ventricular cavity volume and ejection fraction for conventional estimation methods and 3d surface fitting. *Journal of the American Heart Association*, 8(6):e009124, 2019. doi: 10.1161/JAHA.118.009124. URL https://www.ahajournals.org/doi/abs/10.1161/JAHA.118.009124.

[36] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A. Cook, Antonio de Marvao, Timothy Dawes, Declan P. O'Regan, and et al. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, Feb 2018. ISSN 1558-254X. doi: 10.1109/tmi.2017.2743464. URL http://dx.doi.org/10.1109/TMI.2017.2743464.

[37] D. Ouyang, B. He, Amirata Ghorbani, M. Lungren, E. Ashley, D. Liang, and James Y. Zou. Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. 2019.

[38] David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis Langlotz, Paul Heidenreich, Robert Harrington, David Liang, Euan Ashley, and James Zou. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580, 04 2020. doi: 10.1038/s41586-020-2145-8.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[40] Patricia A. Pellikka, Lilin She, Thomas A. Holly, Grace Lin, Padmini Varadarajan, Ramdas G. Pai, Robert O. Bonow, Gerald M. Pohost, Julio A. Panza, Daniel S. Berman, David L. Prior, Federico M. Asch, Salvador Borges-Neto, Paul Grayburn, Hussein R. Al-Khalidi, Karol Miszalski-Jamka, Patrice Desvigne-Nickens, Kerry L. Lee, Eric J. Velazquez, and Jae K. Oh. Variability in Ejection Fraction Measured By Echocardiography, Gated Single-Photon Emission Computed Tomography, and Cardiac Magnetic Resonance in Patients With Coronary Artery Disease and Left Ventricular Dysfunction. *JAMA Network Open*, 1(4):e181456–e181456, 08 2018. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2018.1456. URL https://doi.org/10.1001/jamanetworkopen.2018.1456.

[41] Joseph Rahman, Emelie Helou, Ramona Gelzer-Bell, Richard Thompson, Chih Kuo, E. Rodriguez, Joshua Hare, Kenneth Baughman, and Edward Kasper. Noninvasive diagnosis of biopsy-proven cardiac amyloidosis. *Journal of the American College of Cardiology*, 43:410–5, 02 2004. doi: 10.1016/j.jacc.2003.08.043.

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

[43] Paolo Severino, Andrea D'Amato, Mariateresa Pucci, Fabio Infusino, Lucia Birtolo, Marco Mariani, Carlo Lavalle, Viviana Maestrini, Massimo Mancone, and Francesco Fedele. Ischemic heart disease and heart failure: Role of coronary ion channels. *International Journal of Molecular Sciences*, 21:3167, 04 2020. doi: 10.3390/ijms21093167.

[44] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. ISSN 2169-3536. doi: 10.1109/access.2021.3086020. URL http://dx.doi.org/10.1109/ACCESS.2021.3086020.

[45] Erik Smistad, Olivier Bernard, Bjørnar Grenne, Lasse Løvstakken, Andreas Ostvik, Ivar Salte, Daniela Melichova, Thuy Mi Nguyen, Kristina Haugaa, Harald Brunvand, Thor Edvardsen, and Sarah Leclerc. Real-time automatic ejection fraction and foreshortening detection using deep learning. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, PP:1–1, 03 2020. doi: 10.1109/TUFFC.2020.2981037.

[46] Tanawut Tantimongcolwat, Thanakorn Naenna, Chartchalerm Isarankura-Na-Ayudhya, M. Embrechts, and Virapong Prachayasittikul. Identification of ischemic heart disease via machine learning analysis on magnetocardiograms. *Computers in biology and medicine*, 38:817–25, 08 2008. doi: 10.1016/j.compbiomed.2008.04.009.

[47] Paaladinesh Thavendiranathan, Zoran Popovic, Scott Flamm, Arun Dahiya, Richard Grimm, and Thomas Marwick. Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography*, 26, 08 2013. doi: 10.1016/j.echo.2013.07.017.

[48] Kristian Thygesen, Joseph S. Alpert, Harvey D. White, and null null. Universal definition of myocardial infarction. *Journal of the American College of Cardiology*, 50(22): 2173–2195, 2007. doi: 10.1016/j.jacc.2007.09.011. URL https://www.jacc.org/doi/abs/10.1016/j.jacc.2007.09.011.

[49] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks, 2014. URL https://arxiv.org/abs/1412.0767.

[50] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A closer look at spatiotemporal convolutions for action recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6450–6459, 2018. doi: 10.1109/CVPR.2018.00675.

[51] Du Tran, Heng Wang, Lorenzo Torresani, and Matt Feiszli. Video classification with channel-separated convolutional networks, 2019. URL https://arxiv.org/abs/1904.02811.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[53] Ola Vedin, Carolyn S.P. Lam, Angela S. Koh, Lina Benson, Tiew Hwa Katherine Teng, Wan Ting Tay, Oscar Ö. Braun, Gianluigi Savarese, Ulf Dahlström, and Lars H. Lund. Significance of ischemic heart disease in patients with heart failure and preserved, midrange, and reduced ejection fraction. *Circulation: Heart Failure*, 10(6):e003875, 2017. doi: 10.1161/CIRCHEARTFAILURE.117.003875. URL https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.117.003875.

[54] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sebastien Ourselin, and et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, Jul 2018. ISSN 1558-254X. doi: 10.1109/tmi.2018.2791721. URL http://dx.doi.org/10.1109/TMI.2018.2791721.

[55] Chaoning Zhang, Philipp Benz, Dawit Mureja Argaw, Seokju Lee, Junsik Kim, Francois Rameau, Jean-Charles Bazin, and In So Kweon. Resnet or densenet? introducing dense shortcuts to resnet. In *2021 IEEE Winter Conference on Applications of*

Computer Vision (WACV), pages 3549–3558, 2021. doi: 10.1109/WACV48630.2021.00359.

[56] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. *UNet++: A Nested U-Net Architecture for Medical Image Segmentation: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, volume 11045, pages 3–11. 09 2018. ISBN 978-3-030-00888-8. doi: 10.1007/978-3-030-00889-5_1.

[57] Douglas P. Zipes, Peter Libby, Rober O. Bonow, Douglas L. Mann, Gordon F. Tomaselli, and Eugene Braunwald. *Braunwald's heart disease: a textbook of cardiovascular medicine*. Elsevier, Philadelphia, PA, 11th ed. edition, 2019.