Article

Semi-Supervised Segmentation of Echocardiography Videos Using Graph Signal Processing

Marwa Chendeb El rai 1,2,*, Muna Darweesh 2,† and Mina Al-Saad 2,†

1 Mathematics Department, School of Arts and Sciences, American University in Dubai, Dubai 28282, United Arab Emirates
2 College of Engineering and Information Technology, University of Dubai, Dubai 14143, United Arab Emirates
* Correspondence: melrai@aud.edu or m_chendeb@hotmail.com; Tel.: +971-50-589-8091
† These authors contributed equally to this work.

Abstract: Machine learning and computer vision algorithms can provide a precise and automated interpretation of medical videos. The segmentation of the left ventricle of echocardiography videos plays an essential role in cardiology for carrying out clinical cardiac diagnosis and monitoring the patient’s condition. Most of the developed deep learning algorithms for video segmentation require an enormous amount of labeled data to generate accurate results. Thus, there is a need to develop new semi-supervised segmentation methods due to the scarcity and costly labeled data. In recent research, semi-supervised learning approaches based on graph signal processing emerged in computer vision due to their ability to avail the geometrical structure of data. Video object segmentation can be considered as a node classification problem. In this paper, we propose a new approach called GraphECV based on the use of graph signal processing for semi-supervised learning of video object segmentation applied for the segmentation of the left ventricle in echocardiography videos. GraphECV includes instance segmentation, extraction of temporal, texture and statistical features to represent the nodes, construction of a graph using $K$-nearest neighbors, graph sampling to embed the graph with small amount of labeled nodes or graph signals, and finally a semi-supervised learning approach based on the minimization of the Sobolov norm of graph signals. The new algorithm is evaluated using two publicly available echocardiography videos, EchoNet-Dynamic and CAMUS datasets. The proposed approach outperforms other state-of-the-art methods under challenging background conditions.

Keywords: echocardiography; video object segmentation; deep learning; graph signal processing; semi-supervised learning

1. Introduction

The World Health Organization (WHO) reports that cardiovascular diseases are the major cause of death with the estimation that 17.9 million people die every year [1]. In the last two decades, advancement in imaging technology and machine learning techniques have advanced diagnosing and treating cardiovascular diseases. Echocardiography is a safe and low-cost test for cardiac diagnosis [2]. It is a non-invasive examination which observes all the structures of the heart, namely the valves and the cavities (atria and ventricles). Echocardiography is able to explore the cardiac origin of symptoms, such as shortness of breath, chest pain, or malaise. It evaluates the impact of a disease, such as high blood pressure or pulmonary arterial hypertension or certain medications, on the heart. Echocardiography diagnoses heart failure which prevents blood from flowing back when it enters one of the heart chambers or if is expelled by the heart.

The left ventricle is the principal pumping cavity of the heart. It pumps blood rich in oxygen into the aorta and to the rest of the body. Most cardiac functions, such as myocardial motion analysis and ejection fraction estimation, are determined from the left ventricle. In the last two decades, advancement in imaging technology, computer vision, and the
machine learning techniques have advanced the diagnosis and treatment of cardiovascular diseases. These modern imaging technologies have improved the diagnostic procedures, which in turn increase accuracy and optimize the workload of healthcare workers. Despite the fact that deep learning contributed to the medical diagnosis, there are many obstacles that need to be resolved before deployment. Specifically, deep learning techniques require huge amounts of data to perform as competently as human judgment. Furthermore, due to privacy laws and various standards applied in different healthcare industries, medical data are rarely available compared to the data obtainability in other research fields [3]. In addition, the labeling process of medical data are considered complex, time-consuming, and require experienced healthcare professionals. Therefore, the main idea of this research is to develop a method called GraphECV that requires little labeled data to be used in the semi-supervised segmentation of the left ventricle. It is based on the use of graph signal processing.

Recently, a semi-supervised Background Subtraction approach proposed by [4] and based on the theory of graph signal processing was applied on video object segmentation (VOS). The latter method was updated to perform the semi-supervised segmentation of synthetic aperture radar (SAR) images, where left ventricular objects have different sizes and are embedded in complex environments [5]. In this work, we propose a semi-supervised learning segmentation algorithm for the echocardiography videos called GraphECV. Our algorithm is inspired by the work of [4,6]. GraphECV has the advantage of seeking less labeled data during the training phase than other deep learning approaches while adapting to complex background and texture of echoardiography scenarios. Experimental findings show the effectiveness of our proposed method; it outperforms the other state-of-the-art approaches.

The principal contributions of this work are as follows:

• Our method supports a new semi-supervised learning model for the left ventricle in echocardiography videos that integrates the graph signal processing where nodes are classified into the left ventricle or background.
• The motion, temporal, statistical, and texture features are used to represent the nodes on the graph. This integration has not appeared in the literature.
• The experiments were applied over two public datasets: EchoNet Dynamic and CAMUS. Despite the scarcity of labeled data, GraphECV surpassed many of the state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 summarizes the related works. Section 3 discusses the basic concepts and the proposed semi-supervised learning segmentation algorithm for echocardiography videos based on Graph Signal Processing GraphECV. Section 4 presents the experimental results including the description of datasets and the analysis of ablation studies. Finally, Section 5 concludes and draws future directions of this research.

2. Related Work

This section briefly surveys: (1) Graph Signal Processing and its application in computer vision, and (2) supervised and semi-supervised Video Object segmentations (VOS).

2.1. Graph Signal Processing

Graph Signal processing (GSP) is a domain of signal processing that deals with data illustrated on graphs. It reflects the interaction between two connected fields: applied mathematics and signal processing. The data are depicted as signals and defined in groups of nodes on a weighted graph. It gives a natural representation that integrates both the data and the underlying structure of the geometry [7]. In video processing, GSP has played an important role in analyzing natural signals that are in irregular domains [8]. It is a very beneficial task as it considers the spatio-temporal relationships of the pixels. GSP and its application in the field of machine learning were widely discussed in [7].
2.2. Video Object Segmentation

Although complex, the methods related to the video object segmentation (VOS) emerged after abundant research tools were used for image object segmentation. VOS can be categorised into three main groups: supervised, semi-supervised and unsupervised learning methods. The VOS was investigated by the classical method such as conditional random field (CRF) [9,10] and by the use of deep learning algorithms where authors explored the embedding of crucial and challenging temporal information in VOS [11–13]. A common choice for reflecting the temporal relation between frames in VOS is optical flow [14]. Segflow proposed by [11] introduced a second channel for optical flow [15] besides the CNN network used for image segmentation. In order to utilize most of the temporal information relevant to VOS, authors in [16] created a TMANET method which integrates the attention mechanism with convolutional neural networks. Another study [17] worked to find the similarity between two consecutive frames to obtain the attention coefficients. To boost more the VOS accuracy, ref. [18] computed the motion cue to strengthen the temporal representation of the target frame and the neighbours. Some works in deep learning tackle the problem of VOS as a classical image segmentation by considering the frames independently. The main drawback is the failure to capture the dynamics of the video [19,20]. In addition to the temporal features, authors used sequential model long short-term memory (LSTM) to learn the interference of redundant video frames [21,22]. The deep learning methods require a large amount of annotated data to obtain high performance levels. Many studies outlined the prediction of labels with few labeled samples in the training phase [23,24]. Proposal-generation, refinement and merging for video object segmentation (PReMVOS) automatically generates pixel masks over video sequences, while the first frame is annotated [25]. Authors in [26] introduced a new approach that extracts texture features for image indexing and retrieval in the biomedical field. Authors in [27] proposed a study to predict COVID-19 for diabetic patients based on a fuzzy interface system and machine learning algorithms.

Echocardiography (cardiac ultrasound imaging) is primarily used as a clinical tool for the evaluation of different cardiovascular functions [28]. The authors of [29] proposed a semi-segmentation algorithm to segment the left ventricle endocardium from echocardiography videos and adaptive spatio-temporal semantic calibration to ensure the alignment of the feature maps for the consecutive frames, which will reduce the effect of the speckle noise. In addition, the acquired temporal information from the feature map of the neighboring frame to the current frame is included to improve the segmentation performance. However, its performance might degrade with irregular cardiac motion or low contrast videos. Another approach proposed by Sultan et al. is to segment the anterior mitral leaflet (AML) which is required to diagnose rheumatic heart disease. In this algorithm, the video frames are converted to virtual M-mode space after specifying a single point initialization on the posterior wall of the aorta and then segmenting the AML in this space. However, the proposed algorithm has low robustness since a single seed information is at hand, which might affect the segmentation process, and the algorithm misses the AML tip. A joint learning approach for the spatio-temporal echocardiographic sequence of the left ventricular motion tracking and segmentation was proposed by Ta et al. The features learned from both the segmentation and motion tracking are joined bi-directionally, thus utilizing the features that might not be captured from one branch. The semi-segmentation technique adopted is U-Net. The proposed technique includes physiological constraints which ensure realistic cardiac functioning, which helps in reducing the dependency on the accurate segmentation. ATTIA et al. proposed an automated segmentation technique that extracts the intra-frame and the motion information [2]. Sigit et al. reported a segmentation method that applies the B-Spline to identify the cardiac cavity in the initial frame and optical flow to track and detect the border for each frame in the video [30].
3. Graph Signal Processing and Semi-Supervised Echocardiography Video Segmentation

Figure 1 displays a graphical illustration of our proposed semi-supervised left ventricle segmentation. The framework proposed in this work can be divided into the following steps: a deep learning algorithm FgSegNet_S [31] to segment the left ventricle; handcrafted features extracted from each instance, including optical flow, motion, texture, and statistical features crucial to illustrate the spatio-temporal information of each instance node; graph construction using K-nearest neighbors; graph signal revealed by the annotated images; sampling of graph signals; and finally, to label all the nodes and reconstruct the graph, a semi-supervised Sobolev norm minimization technique is applied. The latter permits us to evaluate the unlabeled nodes and classifies them between the left ventricle and the background with a limited input amount of labeled samples. The sampling known in classical digital signal processing is identical to the sampling of signals or labeled data on the graphs. The concept used here is to reconstruct all the signals from the embedded labeled data belonging to the left ventricle.

Figure 1. The pipeline of the proposed GraphECV method applies instance segmentation. The motion, temporal, textural, and statistical features have been extracted to represent the nodes on the graph. The graph signal propagates the labeled data into the graph. It is used to identify if a node is a left ventricle or a background. A set of nodes are sampled via graph sampling. The reconstruction of all the labels in the graph was achieved by the use of a semi-supervised learning algorithm.

3.1. Introduction to Signal Processing on Graphs

GSP is an expanding area of research that extends classical analytical methods to non-regular fields, exploiting the topology of the underlying graphs relying on the Laplacian and Fourier analyses. GSP is therefore a crucial tool, having the ability to associate signals (left ventricle activity) with the other parts of the heart considered as background in this study. GSP can be defined as an undirected graph \( G = (\nu, \eta) \) with a collection of nodes and a collection of edges \( \eta = (i, j) \) having \( w_{ij} \) as weight reflecting the similarity and correlation between the samples at nodes \( i \) and \( j \); \( w_{ij} \) is an element of the adjacency matrix \( W \); \( w_{ij} = 0 \) when \( (i, j) \notin \eta \). The graph signal emerging over the nodes of graph \( G \) is defined as \( f \in \mathbb{R}^N \). A subgroup \( A \) consists of \( M \) sampled data, where the ratio of \( M \) to \( N \) is known as the sampling density. Hence, the sampled graph signal is defined as \( y(A) = Sy \), where
\( S = [\delta_{a1}, \delta_{a2}, \ldots, \delta_{aN}]^T \) is an \( N \) dimension Kronecker column vector [32] acting as binary decimation matrix.

3.2. Instance Segmentation

The complexity nature, gray color, and background of echocardiography images is different from optical images. They require an additional module in any segmentation method to enhance the contrast between the features in order to differentiate between the different parts of an ultrasound image. In 2018, the FgSegNet_S method [31] was developed for video foreground background segmentation. It includes a Feature Pooling Module (FPM) integrated between the encoder and the decoder CNN networks of the SegNet [33] segmentation approach. The FPM has the ability to elicit multi-scale features from the input encoder CNN. The extracted features exemplify the input to the following decoder CNN. In addition, FPM guarantees a strong feature pooling versus any probe motion [34]. In addition, FPM grants better classification of the uncertainty sourced by speckle noise [35]. An ablation study was conducted to compare many semantic and instance segmentation methods. For echocardiography videos, the performance using FgSegnet_S outstripped the other segmentation methods. The results are shown in Table 1.

Table 1. Dice Coefficient score of our proposed method when using different segmentation methods for nodes representation. The percentage of annotated data is 5%.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Unet</th>
<th>DeepLab</th>
<th>Mask-RCNN</th>
<th>FgSegNet_S</th>
</tr>
</thead>
<tbody>
<tr>
<td>EchoNet-Dynamic</td>
<td>0.6432</td>
<td>0.7890</td>
<td>0.742</td>
<td>0.9113</td>
</tr>
<tr>
<td>CAMUS</td>
<td>0.6995</td>
<td>0.7890</td>
<td>0.7695</td>
<td>0.9270</td>
</tr>
</tbody>
</table>

3.3. Feature Extraction and Nodes Representation

Many features of interest can be integrated, such as texture, motion and optical flow features pivotal in medical videos. Thus, the temporal information was included by the estimation of optical flow [14]. The idea is to use the echo features which are different from optical parameters. With gray scale and no color information, echocardiography videos acquired using ultrasound imaging system vary glaringly from optical images. It is essential to determine features that characterize the left ventricular from the other tissues and background. Statistical features are estimated based on the fact that ultrasound tissue can be characterized by a generalized Gamma distribution (Generalized Gamma Distribution) [36]. This distribution can help distinguish between the left ventricle, other tissues, and background, as the left ventricular pixels appear darker than the others.

3.3.1. Statistical Representation of Echocardiography Data

Echocardiography videos, like any other ultrasound images, are depicted by the presence of a salt and pepper pattern called speckles. They are the result of an important spatial heterogeneity between close pixels. The study of backscattered echos from tissues demands a proper analysis of the ultrasonic signals which can be provided through their statistical description [37]. The parameters of a statistical model permit the generation of discriminative descriptors that are crucial for the classification and identification of left ventricle and other tissues [38]. The experiments depicted in [36] show that the generalized Gamma distribution (GTD) can precisely describe the behavior of myocardial tissue and left ventricle in echocardiography images. The parameters of (GTD) offer a posterior probability which is helpful for classification and segmentation [36,39]. Hence, the gray level statistics of echocardiography videos can be modeled by (GTD) [36,40] having the following probability density function (PDF):

\[
p(x) = \frac{\phi \kappa^\kappa}{\theta \Gamma(\kappa)} \left( \frac{x}{\theta} \right)^{\kappa - 1} \exp\left( -\kappa \left( \frac{x}{\theta} \right)^\kappa \right). \tag{1}
\]
In this equation, \( \Gamma(\cdot) \) symbolizes the Gamma function. The scale \( \theta \), the shape \( \kappa \), and power’s distribution \( \varphi \) parameters are estimated using Mellin Transform and second-kind statistics [36,41]. The log-cumulant expressions of \( (G^\Gamma D) \) can be expressed as follows [42]:

\[
\hat{\zeta}_1 = \log(\theta) + \frac{\Phi_0(\kappa) - \log(\kappa)}{\varphi}
\]

\[
\hat{\zeta}_i = \frac{\Phi_0(i-1, \kappa)}{\varphi^i} \quad i = 1, 2, 3, ...
\]

where \( \Phi_0(x, y) \) and \( \Phi_0(x) \) are the Polygamma and Diagamma functions, respectively. The higher order of Polygamma function leads to the estimation of the shape parameter as follows:

\[
\hat{\kappa} = \frac{\lambda_1^3}{3\lambda_0^2} + 3\left[\frac{-O}{2} + \sqrt{\frac{O^2}{4} + \frac{T^3}{27}}\right] + 3\left[\frac{-O}{2} + \sqrt{\frac{O^2}{4} + \frac{T^3}{27}}\right]
\]

where \( O \), \( T \) quantities are defined as:

\[
O = \frac{3\lambda_0\lambda_2 - \lambda_1^2}{3\lambda_0^2}
\]

\[
T = \frac{3\lambda_1^3 - 9\lambda_0\lambda_1\lambda_2 + 27\lambda_0^2\lambda_3}{27\lambda_0^3}
\]

\( \lambda_i \) are expressed as follows:

\[
\lambda_0 = 8\hat{\zeta}_3^2
\]

\[
\lambda_1 = 4(3\hat{\zeta}_3^2 - 2\hat{\zeta}_2^2)
\]

\[
\lambda_2 = 2(3\hat{\zeta}_3^2 - 8\hat{\zeta}_2^2)
\]

\[
\lambda_3 = \hat{\zeta}_3^2 - 8\hat{\zeta}_2^2
\]

Using the log-cumulants calculated in Equations (2) and (3) and the estimated \( \kappa \) in Equation (4) permits us to evaluate the parameters \( \varphi \) and \( \theta \) as:

\[
\hat{\varphi} = \text{sign}(-\hat{\zeta}_3)\sqrt{\frac{1}{\hat{\zeta}_2}\Phi_0(1, \hat{\kappa})}
\]

\[
\hat{\theta} = \exp\left\{\hat{\zeta}_1 - \frac{\Phi_0(\hat{\kappa}) - \log(\hat{\kappa})}{\hat{\varphi}}\right\}
\]

The classical statistical features, such as kurtosis, skewness, mean, and standard deviation, are extracted too.

3.3.2. Texture Features

In general, the texture features are used due to their major influence in image segmentation [43]. In the case of echocardiography video frames having complex separations between the boundary regions, the texture features are essential to discriminate between boundaries as they are considered a function of the spatial variation of pixel intensities of gray values. Consequently, local binary pattern (LBP), entropy, and intensities are determined to represent the texture features [44]. LBP texture descriptors have a potential capacity to distinguish between the tiny differences in terms of topography and texture [44] as they contain details from different areas of the left ventricle [45]. The entropy is determined as the first occurrence for texture analysis.
3.3.3. Nodes Representation of Segmented Instances

The first step of the framework generates output masks by applying FgSegNet_S. For each instance, the motion, temporal, statistical, and texture features are estimated. Then, they are concatenated to represent the instances on the vertices of the graph. The features constitute dimensional vectors of length 148.

3.4. Graph Construction

The GraphECV algorithm uses K-nearest neighbors (k-NN) reported in most of the STOA methods for graph construction. We consider $X = [x_1, x_2, ..., x_N]^T$ the matrix of features of $N$ vertices. By linking $K$ neighbors of each vertex or node, the following kernel is implemented to estimate the weight of each edge:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

(10)

where $w_{ij}$ is calculated using a Gaussian kernel [46] which reflects the similarity and correlation between the samples at nodes $i$ and $j$. The high values of $w_{ij}$ indicate that the instances are well correlated; $\sigma$ is the standard deviation expressed as:

$$\sigma = \frac{1}{|\epsilon + N|} \sum_{(i,j)\in\epsilon} \|x_i - x_j\|^2_2$$

(11)

3.5. Graph Signals

In this work, the matrix $Y \in \{0,1\}^{N \times 2}$ is considered a graph signal with two classes ($p = 2$): left ventricular and background. Each row of $Y$ reflects the segmented region belonging to the left ventricular ($[0,1]$) or to the background ($[1,0]$). In order to identify whether the node is background or a left ventricle, the intersection over union and the intersection over vertex are computed [4].

3.6. Semi-Supervised Learning

The GraphECV algorithm requires the Sobolev norm which is a semi-supervised learning approach to construct the graph after the sampling operation. Variational splines or combinatorial Laplace operator are introduced as tools to minimize the Sobolev norms [47]. The graph signal of labeled (or sampled) data is defined as:

$$y_p(A) = Sy_p$$

(12)

where $p = 1, 2$. The Sobolev norm of the graph signals is expressed as:

$$\|z_p\|_{\alpha,\epsilon} = \|(L + \epsilon I)\alpha/2 z_p\|_2^2$$

(13)

Consequently, the semi-supervised learning approach aiming to minimize the Sobolev norm can be watched as an optimisation problem expressed as:

$$z_p z_p^T (L + \epsilon I)^\alpha z_p \quad s.t \quad S z_p - y_p(A) = 0$$

(14)

where $L$ is the combinatorial Laplacian matrix of the graph $G$, and $I$ is the identity matrix. As $(L + \epsilon I)$ is an invertible matrix for $\epsilon > 0$ in non-directed graphs [48], the solution of the optimization problem in Equation (14) can be revealed as:

$$\hat{Z} = ((L + \epsilon I)^{-1})^a S^T (S((L + \epsilon I)^{-1})^a S^T)^{-1} y(A).$$

(15)

Experimentally, $\epsilon$ is set to 0.2, and $a$ to 1 in this work. These values provide the best results after running different experiments for $\epsilon = [0.2, 0.1, 10, 20]$ and $a = [1, 2]$. The graph signal processing toolbox developed by [49] is employed to solve the optimization problem introduced in Equation (14).
4. Experimental Results

This section introduces the echocardiography videos datasets used in the current work, the evaluation metrics applied to assess the performance of the proposed methodology, the implementation details, and the experiments performed during the implementation of our proposed method GraphECV.

4.1. Datasets

4.1.1. Echonet-Dynamic Dataset

The EchoNet-Dynamic Dataset [50] is the first dataset reported in the literature with 10,030 echocardiography videos. The 2-D gray scale videos are generated from 10,030 individuals with unique visits. The videos have a resolution of 112 × 112 pixels, and it is a four-chamber view [50]. Two frames of each video were manually labeled by medical professionals [29].

4.1.2. CAMUS Dataset

The second dataset adopted for this research is the Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) [51]. It was introduced to the research community in 2019. This 2-D dataset contains the medical exams of 500 patients. The data were collected at various acquisition settings with no prerequisite, so some cases were challenging to trace. In addition, in some cases, the wall is not visible. A portion of the data were acquired in five-chambers view settings rather four-chambers view setting, since the probe orientation was unfeasible. This would produce realistic scenarios [51].

4.2. Evaluation Metric

The evaluation criteria is the Dice coefficient (DC) or $F_1$ measure. The Dice coefficient is a geometric metric which measures the pixel similarities between ground truth data and their corresponding predicted segmentation. It is expressed as follows [52]:

$$DC = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN}$$

(16)

where $X$ is the predicted data, and $Y$ is the ground truth. The Dice coefficient determines the overlap between $X$ and $Y$. $TP(\text{TruePositive})$, $FP(\text{FalsePositive})$, and $FN(\text{FalseNegative})$ represent the amount of pixels which are correctly assigned as labels, incorrectly assigned as labels, and incorrectly assigned as no labels in $X$, respectively.

4.3. Implementation Details

Python 3.7 and Detectron [53] were used for the implementation of instance and semantic segmentation. FgSegNet_S was trained for 200 epochs using a learning rate of 0.00035 and a batch size of 5. The graph signal processing toolbox [49] was utilized for the reconstruction of graph signals. The experiments were implemented on a powerful NVIDIA Geforce RTX GPU.

4.4. Results

Several ablation studies were conducted over the datasets to analyze the performance of our proposed model (GraphECV) which implicates many parameters such as the percentage of labeled ground truth frames used during the training process, the number of $k$ neighbors for $k$-NN in the graph construction, and the parameters $\alpha$ and $\epsilon$ for the Sobolev norm. Ablation studies were carried out for these parameters. In addition, experimental results analyze and discuss some components of the framework in Figure 1 such as the segmentation method. To represent the nodes on the graph during the graph construction, many semantic segmentation methods were applied. FgSegNet_S attained the best performance due to the FPM module implemented between the encoder and decoder. FPM was able to better segment the dense spatial of echocardiography videos. FgSegNet_S outperformed the STOA segmentation algorithms such as Mask-RCNN [54], Unet [55]
and Deeplab [56] especially in the case of a small amount of annotated data (5%). Table 1 depicts the segmentation’s performance in terms of the Dice coefficient score for different segmentation methods. K-NN is responsible of the construction of the graph. Table 2 briefs the performance of GraphECV with various values of $k$ parameter ($k = 5$, $k = 10$, $k = 20$, and $k = 30$). For both datasets, the best results were calculated for $k = 30$ where all the nodes were connected. On the other hand, for small $k$ values, the graph was mislaying global information of the database. For this experiment, 5% of annotated data was used.

**Table 2.** Dice coefficient score of our proposed method with change in the construction of the graph. This ablation encompasses $K$-nearest neighbors with $k = 5$, $k = 10$, $k = 20$, and $k = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$k = 5$</th>
<th>$k = 10$</th>
<th>$k = 20$</th>
<th>$k = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EchoNet-Dynamic</td>
<td>0.8320</td>
<td>0.8523</td>
<td>0.8789</td>
<td>0.9113</td>
</tr>
<tr>
<td>CAMUS</td>
<td>0.8459</td>
<td>0.8648</td>
<td>0.8837</td>
<td>0.9270</td>
</tr>
</tbody>
</table>

The parameters $\alpha$ and $\epsilon$ of the semi-supervised learning block were associated to the Sobolev minimization. Sobolev minimization experiments were performed for $\epsilon = 0.2$, 0.5, 1, 20, and $\alpha = 1, 2$. Tables 3 and 4 summarize the performance of the minimization of the Sobolev for $\alpha = 1$ and $\alpha = 2$, respectively. The best results were obtained for $\epsilon = 0.2$ and $\alpha = 1$. Higher values of $\alpha$ make the Laplacian matrix denser, which at the same time result in computational and memory problems.

**Table 3.** Dice coefficient score of our proposed method with ablation encompassing the Sobolev minimization parameters with $\alpha = 1$. The percentage of annotated data is 5%.

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon = 0.1$</th>
<th>$\epsilon = 0.2$</th>
<th>$\epsilon = 10$</th>
<th>$\epsilon = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EchoNet-Dynamic</td>
<td>0.8589</td>
<td>0.9113</td>
<td>0.8954</td>
<td>0.8561</td>
</tr>
<tr>
<td>CAMUS</td>
<td>0.8796</td>
<td>0.9270</td>
<td>0.9014</td>
<td>0.8924</td>
</tr>
</tbody>
</table>

**Table 4.** Dice coefficient score of our proposed method with ablation encompassing the Sobolev minimization parameters with $\alpha = 2$. The percentage of annotated data is 5%.

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon = 0.1$</th>
<th>$\epsilon = 0.2$</th>
<th>$\epsilon = 10$</th>
<th>$\epsilon = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EchoNet-Dynamic</td>
<td>0.8321</td>
<td>0.8652</td>
<td>0.8789</td>
<td>0.8958</td>
</tr>
<tr>
<td>CAMUS</td>
<td>0.8591</td>
<td>0.8687</td>
<td>0.8956</td>
<td>0.9087</td>
</tr>
</tbody>
</table>

For the variation of percentage of labeled data, the experiments were conducted using the percentages of [5%, 10%, 20%, 30%, and 50%]. The nodes of the graph were weighted upon calculation of the motion, temporal, texture, and statistical features. To show the discriminative effect of the statistical parameters of $GT_D$, GraphECV was applied without the integration of statistical features. The results obtained in both cases are reported in Table 5. While the framework was trained using 5% and 30% of labeled data, it was proven that our solution to integrate the statistical parameters of $GT_D$ outstripped by approximately 10% the case where the $GT_D$ parameters were not included.

**Table 5.** Average Dice coefficient reported on the datasets without and with the integration of statistical parameters of $GT_D$; 5%, and 30% of labeled data were used during the training process.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>With $GT_D$</th>
<th>Without $GT_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>30%</td>
</tr>
<tr>
<td>EchoNet-Dynamic</td>
<td>0.9113</td>
<td>0.9301</td>
</tr>
<tr>
<td>CAMUS</td>
<td>0.9270</td>
<td>0.9329</td>
</tr>
</tbody>
</table>

Our method was compared with the semi-supervised and supervised STOA methods for VOS. The supervised STOA methods include proposal-generation, refinement and
merging for video object segmentation (PReMVOS) [57] and one shot video object segmentation (OSVOS) [58], while the semi-supervised STOA methods are temporal memory attention network (TMANet) [16] and a corrective fusion network for efficient semantic segmentation on video (Accel) [14]. For fair comparison, these methods were applied on the same test datasets. Figures 2 and 3 show the visual results of the proposed GraphECV method and the STOA algorithms for the EchoNet-Dynamic and CAMUS datasets, respectively. Tables 6 and 7 display the comparisons of the qualitative results of the GraphECV approach on EchoNet-Dynamic and CAMUS datasets for VOS with the STOA methods spanning all the percentage of labeled data. The whole training data (100%) was used in the case of the EchoNet-Dynamic dataset to compare with the baseline deep learning segmentation method EchoNet-Dynamic developed by [50] (Echone-Dynamic method is called Echonet here to differentiate between the method and the dataset). The performance of our proposed framework surpasses the other STOA methods for all the percentage of labeled data on both datasets. We can observe an improvement of the Dice coefficient when the percentage of labeled data increased from 5% to higher percentages. Although in the case of a very small amount of annotated data (5%), our results show competitive performance compared to other STOA methods trained over 50% or fully annotated data. This is mainly due to the semi-supervised learning yielding rigorous discrimination of the left ventricle on graph nodes.

<table>
<thead>
<tr>
<th>Original</th>
<th>GT</th>
<th>PreMVOS</th>
<th>Tmanet</th>
<th>Accel</th>
<th>Echonet</th>
<th>OSVOS</th>
<th>Ours</th>
</tr>
</thead>
</table>

![Figure 2. EchoNet-Dynamic dataset: visual results of the proposed GraphECV method and other STOA algorithms. Different frames from different videos with many motions of the left ventricle are presented from top to bottom. From left to right: original frame, the ground truth, PreMVOS [57], Tmanet [16], Accel [14], Echonet [50], OSVOS [58], and our proposed GraphECV.](image)
Figure 3. CAMUS dataset: visual results of the proposed GraphECV method and other STOA algorithms. Different frames from different videos with many motions of the left ventricle are presented from top to bottom. From left to right: original frame, the ground truth, PreMVOS [57], TMANET [16], Accel [14], OSVOS [58], and our proposed GraphECV.

Table 6. EchoNet-Dynamic dataset: Statistical comparison based on the average Dice coefficient with supervised and semi-supervised STOA methods with different percentage of labeled training data.

<table>
<thead>
<tr>
<th>Method</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>50%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSVOS</td>
<td>0.8012</td>
<td>0.8745</td>
<td>0.8823</td>
<td>0.8912</td>
<td>0.9025</td>
<td>0.9132</td>
</tr>
<tr>
<td>PreMVOS</td>
<td>0.7532</td>
<td>0.7756</td>
<td>0.7845</td>
<td>0.7960</td>
<td>0.8023</td>
<td>0.8245</td>
</tr>
<tr>
<td>Echonet</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9200</td>
</tr>
<tr>
<td>Accel</td>
<td>0.8496</td>
<td>0.8579</td>
<td>0.8598</td>
<td>0.8609</td>
<td>0.8641</td>
<td>0.8756</td>
</tr>
<tr>
<td>TMANET</td>
<td>0.8523</td>
<td>0.8699</td>
<td>0.8752</td>
<td>0.8895</td>
<td>0.9027</td>
<td>0.9132</td>
</tr>
<tr>
<td>Ours</td>
<td>0.9113</td>
<td>0.9209</td>
<td>0.9285</td>
<td>0.9301</td>
<td>0.9355</td>
<td>0.9389</td>
</tr>
</tbody>
</table>

Table 7. CAMUS dataset: Statistical comparison based on the average Dice coefficient with supervised and semi-supervised STOA methods with different percentage of labeled training data.

<table>
<thead>
<tr>
<th>Method</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSVOS</td>
<td>0.8352</td>
<td>0.8401</td>
<td>0.8479</td>
<td>0.8654</td>
<td>0.8845</td>
</tr>
<tr>
<td>PreMVOS</td>
<td>0.8629</td>
<td>0.8710</td>
<td>0.8746</td>
<td>0.8810</td>
<td>0.89553</td>
</tr>
<tr>
<td>Accel</td>
<td>0.8745</td>
<td>0.8891</td>
<td>0.8954</td>
<td>0.9018</td>
<td>0.9132</td>
</tr>
<tr>
<td>TMANET</td>
<td>0.8862</td>
<td>0.8954</td>
<td>0.9018</td>
<td>0.9132</td>
<td>0.9258</td>
</tr>
<tr>
<td>Ours</td>
<td>0.9270</td>
<td>0.9257</td>
<td>0.9289</td>
<td>0.9329</td>
<td>0.9396</td>
</tr>
</tbody>
</table>

5. Conclusions

Accurate interpretation and analysis of the echocardiography videos are important in assessing cardiovascular diseases. In this research paper, we suggested a new tool, GraphECV, of semi-supervised learning for echocardiography video segmentation aiming to detect the left ventricle. The framework of GraphECV requires segmentation and extraction of texture, statistical, and temporal features to represent the nodes on the graph, application of K-nearest neighbors to construct the graph, graph sampling by embedding the graph with few labeled data, and at the end, semi-supervised learning to reconstruct the graph. The proposed algorithm was evaluated on two publicly available echocardiography datasets. Through the experiments, GraphECV consistently outstripped several STOA methods by a significant margin.

For future research directions, we intend to address the problem of real-time processing for echocardiography videos which improves the diagnostic process and the healthcare of
the patient. Furthermore, the semi-supervised learning based on Graph Signal Processing can explore other relevant features capable of enhancing the representation of nodes on the graph.

Author Contributions: Methodology, M.C.E.r.; Writing—original draft, M.D.; Writing—review & editing, M.A.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.


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

Abbreviations

The following abbreviations are used in this manuscript:

- **CAMUS** Cardiac Acquisitions for Multi structure Ultrasound Segmentation
- **FPM** Feature Pooling Module
- **GTD** Generalized Gama Distribution
- **GSP** Graph Signal Processing
- **LBP** Local Binary Pattern
- **STOA** State-Of-The-Art
- **VOS** Video Object Segmentation

References


