Automatic segmentation of the left ventricle from pediatric echocardiography images using SegFormer architecture

Melisa Mateu[1], Jimena Olveres-Montiel[2], Boris Escalante-Ramírez[2], Marie Josée Raboisson[3], Joaquim Miró[3], Luc Duong[1]

[1] Department of software and IT engineering, École de Technologie Supérieure, Montreal, Canada
[2] Department of electrical engineering, Universidad Nacional Autónoma de México, Mexico City, Mexico.
[3] Department of cardiology, Centre Hospitalier Universitaire Sainte – Justine, Montreal, Canada

Abstract— **Echocardiography is the most widely used imaging technique for congenital heart disease (CHD) detection, assessing risk, and guiding treatment strategies in pediatric cardiology. However, interpreting and analyzing these types of images can be challenging due to their complexity, which is some cases leads to inter-observer variability. This research work aims to develop an automated left ventricle (LV) segmentation method for pediatric echocardiography images using a semantic transformer model known as SegFormer, for aiding in the measurement of clinical image technique. Semantic transformers have demonstrated exceptional performance in segmentation tasks in recent years, making them a suitable choice for this application. To achieve accurate LV segmentation, the SegFormer model is trained using the EchoNet-Peds dataset, which consists of annotated pediatric echocardiography videos. The experimental results include segmented left ventricle images, evaluated in accuracy, mean absolute error (MAE), recall and dice score metrics for performance comparison with other pediatric segmentation method.**

Keywords— **Echocardiography, segmentation, left ventricle, SegFormer.**

INTRODUCTION

Congenital heart disease (CHD), affects approximately 40,000 births just in the USA per year and about 1 in 4 is critical and will need surgery [1]. According to [2], congenital heart disease can be defined as *"a gross structural abnormality of the heart or intrathoracic great vessels that is actually or potentially of functional significance"*. The CHD are leading cause of infant mortality and can result in chronic disabilities or morbidities [3], [4].

For detecting congenital heart disease in pediatric patients, cardiologists use echocardiography, a non-invasive imaging technique used for heart analysis [5], [6]. In an echocardiography study, abnormalities in the left ventricle (LV) suggest the presence of a cardiopathy, this is why for cardiologists, echocardiography is a very important clinical technique to assess cardiac functions by measuring the heart's ejection fraction or the volumes at the end of systole (ES) and end of diastole (ED) in the cardiac cycle. To achieve these quantifications, the heart's analysis is based on the LV, which is responsible for the blood supply to the body. However, echocardiographic images can be complex to analyze [7], due to the low-quality images with low-contrast regions and speckle noise, in comparison to other medical imagining techniques. These disadvantages often affect interobserver variability, which can negatively impact the accuracy and efficiency of the diagnostic process by increasing subjectivity and reducing the inter-reader reliability [8], [9].

For these reasons, this work proposes the development of a segmentation method that uses pediatric echocardiography and the novel SegFormer architecture, to accurately segment the left ventricle in pediatric patients. This as part of a project to developed a reliable computational system capable of classifying CHD to aid in the processing of echocardiography by offering a tool to reduce the interobserver variability in children.

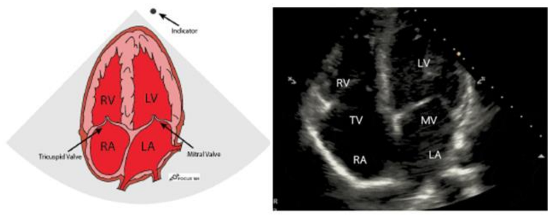
The goal of the present work, is that by using a newly semantic transformer technique, achieve a pediatric left ventricle segmentation performance comparable to the EchoNet-Pediatric, the only computational pediatric segmentation method currently available, which reports a mean dice score of 0.89 for two different cardiac views.

METHODOLOGY

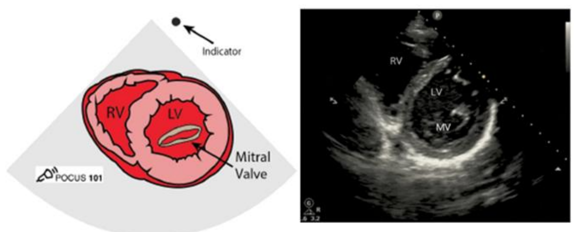
<u>Dataset</u>

We used the publicly available EchoNet-Peds dataset [10], acquired as part of routine clinical care at Lucile Packard Children's Hospital at Stanford University between 2014 and 2021, and published in 2022. This dataset contains echocardiography videos and their human expert annotations.

EchoNet-Peds includes 3,176 videos of apical four-chamber (A4C) view (Fig. 1A), and 4,424 parasternal short axis (PSAX) view videos (Fig. 1B), both views of patients ranging from 0 - 18 years, and the human expert annotations that includes the ejection fraction measurements, and tracings of the left ventricle of these echocardiography videos.



A. Apical four-chamber view, illustration (left) and, ultrasound image (right). Where RV, right ventricle; RA, right atrium; LA, left atrium; LV left ventricle; MV mitral valve; TV, tricuspid valve.



B. Parasternal short – axis view, illustration (left) and, ultrasound image (right). Where RV, right ventricle; LV left ventricle; MV mitral valve.

Figure 1: Cardiac ultrasound different views, illustration and image examples. Source: [11]

SegFormer architecture

SegFormer is a state-of-the-art transformer framework for semantic segmentation, designed to balance efficiency, accuracy, and robustness.

Semantic segmentation extends image classification by moving from classifying entire images to assigning labels at the pixel level. This type of technique has demonstrated exceptional performance in segmentation tasks in recent years, making them a suitable choice for this work application.

SegFormer is an efficient and powerful semantic segmentation architecture, illustrated in Fig. 2, that integrates transformers with lightweight multilayer perceptron (MLP) decoders.

This architecture stands out for two keys features 1) SegFormer employs a novel hierarchically structured Transformer encoder that generates multiscale features. Unlike traditional models, it does not rely on positional encoding, eliminating the need for interpolating positional codes. This prevents performance degradation when the testing resolution differs from the training resolution. 2) Instead of using complex decoders, SegFormer utilizes an MLP decoder that aggregates information from various layers. This design effectively combines local and global attention, resulting in robust feature representations and improving segmentation performance [12], [13], [14].

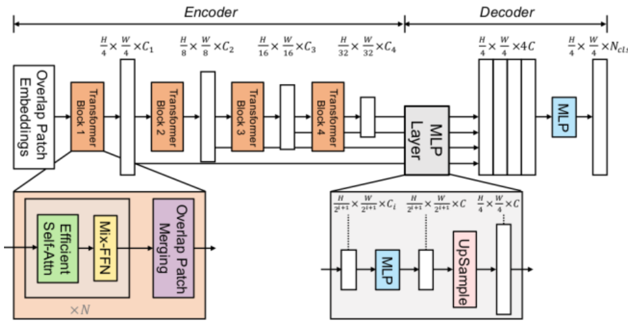


Figure 2: SegFormer architecture consists of two main modules 1) a hierarchical transformer encoder to generate high-resolution crude features and low-resolution fine features; and 2) a lightweight all-MLP decoder to fuse these multi-level features to produce the final semantic segmentation mask. Source:[12]

Segmentation of the left ventricle

For segmenting the LV pediatric patients' echocardiography, the frames at the end of systole and at the end of diastole were extracted from each cardiac ultrasound video. Also, the binary mask for each one of these frames was created by using the annotations included in the EchoNet-Peds database.

The SegFormer was trained using a dataset of the previous extracted frames and their masks, resized to 256x256. The complete dataset was divided in 80%, 15%, 5% for train, test and validation and the Adam optimizer was used with a learning rate of 6E-7 in a 30 epochs training.

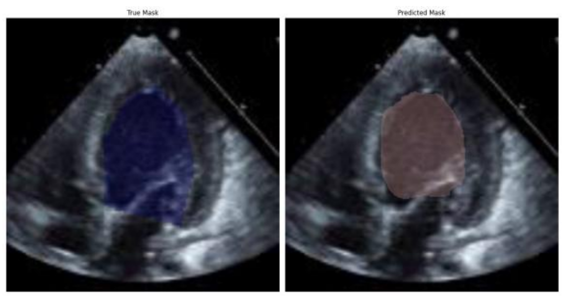
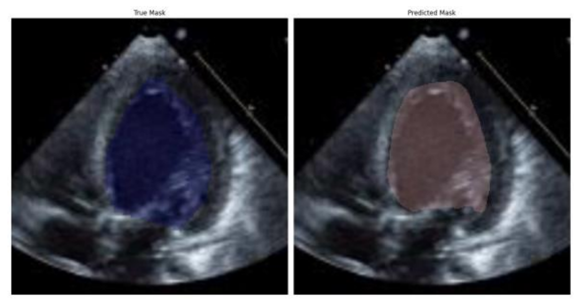
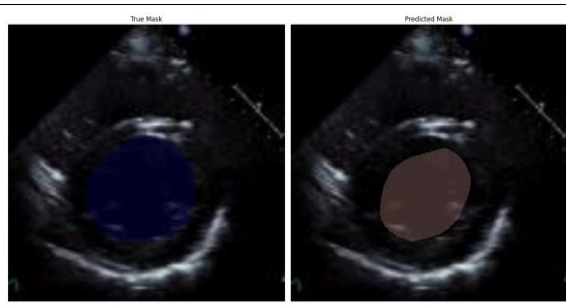
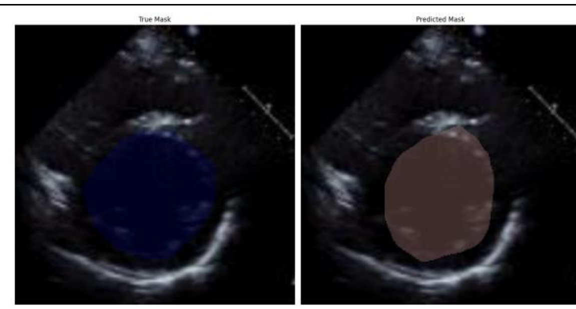
One model was trained for the end of the systole and another model was trained for the end of the diastole for each cardiac view included in the dataset (A4C and PSAX).

The results of the SegFormer for the segmentation task of each model were evaluated for accuracy, dice similarity, recall and MAE, based on the annotated masks from the human expert and the ones generated from the SegFormer.

RESULTS

The SegFormer trained for this pediatric LV segmentation task has as output an estimated binary mask of the left ventricle, for each view (A4C and PSAX) and each part of the cardiac cycle (ED and ES). In Table 1, an example of the results obtained from the model is shown.

Table 1: SegFormer left ventricle's pediatric output mask (pink) compare to the expert's annotated mask (blue)

View	Cardiac cycle moment	
	End of Systole	***End of Diastole***
Apical four chamber (A4C)		
Parasternal Short-Axis (PSAX)		

The mask obtained from the SegFormer training (pink) is compared with the expert's annotated mask (blue). Additionally, both masks are overlapped on the original cardiac ultrasound frame for visual perspective. As well, in Fig 3, an overlapping between both masks and the image is shown.

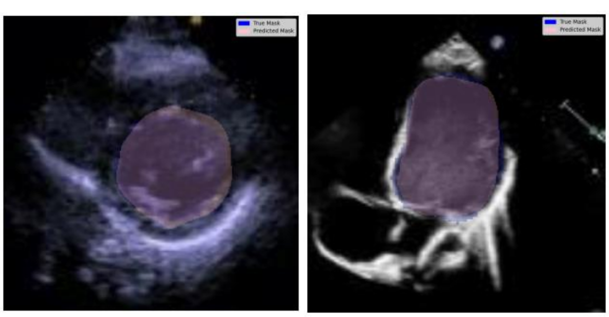


Figure 3: SegFormer output mask (pink) overlapped on its corresponding annotated mask (blue) and cardiac ultrasound image. Diastole PSAX view (left) and diastole A4C view (right).

It can be noticeable that the masks predicted from the segmentation model and the ones previously annotated from the experts show similarities in shape for every case. However, the masks are not equal, while the ones annotated have a more geometrical shape, the predicted masks tend to have a more anatomical shape, showing curves more similar to the LV shape observed directly from the echocardiography.

In Table 2, the metrics' results are shown for both views, at the end of diastole and end of systole. Also, the mean value of each metric is included to provide a more general idea of the results obtained.

It can be observed that the model has a high accuracy for both views, being slightly higher for the PSAX ones. However, this pattern is the opposite for the rest of the metrics, where the A4C view results are higher, or in the MAE case lower, in comparison with the PSAX view. Although, the metrics for both views showed good results.

Table 2: Metrics evaluated on the SegFormer performance for LV segmentation of pediatric patients at ED - ES

Metrics	**SegFormer**	
	A4C view	***PSAX view***
Accuracy	0.967 (0.966 – 0.969)	0.969 (0.970 – 0.968)
Dice score	0.855 (0.880 – 0.831)	0.846 (0.888 – 0.804)
Recall	0.815 (0.874 – 0.756)	0.794 (0.863 – 0.725)
MAE	0.031 (0.033 – 0.030)	0.030 (0.029 – 0.032)

In Table 3, a comparison between the Dice Score of the SegFormer model and EchoNet-Pediatric model is done. Both models were trained in segmenting the pediatric LV and used the same EchoNet-Peds dataset.

Table 3: Dice score comparison between left ventricle's pediatric segmentation models at ED - ES

Model	**Dice Score**	
	A4C view	***PSAX view***
EchoNet-Pediatric	0.891 (0.889 – 0.894)	0.896 (0.894 – 0.898)
SegFormer	0.855 (0.880 – 0.831)	0.846 (0.888 – 0.804)

The comparison is done using the Dice score results at the ED and ES, and the mean value between them. It can be seen that the results of the SegFormer segmentation are comparable to the EchoNet-Pediatric model, both are in the 0.8 order. However, the SegFormer performance on this metric showed to be lower than the results obtained by the EchoNet-Pediatric for both cardiac ultrasound views.

CONCLUSION

The SegFormer architecture accomplished the segmentation task with very high metrics of accuracy and Dice score, and lower values of MAE. The masks obtained from this model present more soft curves, similar to the ones shown on the left ventricle echocardiography images. However, in comparison with EchoNet-Pediatric model, SegFormer showed a lower Dice score value for each view of the pediatric echocardiography.

The semantic Transformer, SegFormer, has proven to be suitable for this medical imaging segmentation task, having good results than can be improved for providing clinicians a novel computational tool that can reduce analysis time and efficiently process large amounts of data, and in a future application support the diagnostic process of congenital heart disease of pediatric patients.

ACKNOWLEDGEMENTS

This research work was supported by NSERC Discovery grant.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] "Data and Statistics on Congenital Heart Defects | CDC." Accessed: Nov. 30, 2023. [Online]. Available: https://www.cdc.gov/ncbddd/heartdefects/data.html#References

[2] S. C. Mitchell, S. B. Korones, and H. W. Berendes, "Congenital Heart Disease in 56,109 Births Incidence and Natural History," *Circulation*, vol. 43, no. 3, pp. 323–332, 1971, doi: 10.1161/01.CIR.43.3.323.

[3] M. Prsa, T. Saroli, A. S. MacKie, A. B. Dancea, J. A. Correa, and M. Asgharian, "Birth prevalence of congenital heart disease," *Epidemiology*, vol. 20, no. 3, pp. 466–468, May 2009, doi: 10.1097/EDE.0B013E31819F3BC0.

[4] M. D. Reller, M. J. Strickland, T. Riehle-Colarusso, W. T. Mahle, and A. Correa, "Prevalence of Congenital Heart Defects in Metropolitan Atlanta, 1998-2005," *J Pediatr*, vol. 153, no. 6, pp. 807–813, Dec. 2008, doi: 10.1016/J.JPEDS.2008.05.059.

[5] A. Ghorbani *et al.*, "Deep learning interpretation of echocardiograms," 2020, doi: 10.1038/s41746-019-0216-8.

[6] D. Ouyang *et al.*, "EchoNet-Dynamic: a Large New Cardiac Motion Video Data Resource for Medical Machine Learning".

[7] C. G. Fonseca *et al.*, "The Cardiac Atlas Project - an imaging database for computational modeling and statistical atlases of the heart," vol. 27, no. 16, pp. 2288–2295, 2011, doi: 10.1093/bioinformatics/btr360.

[8] Y. Lei *et al.*, "Echocardiographic image multi-structure segmentation using Cardiac-SegNet," 2021, doi: 10.1002/mp.14818.

[9] F. Liu, K. Wang, D. Liu, X. Yang, and J. Tian, "Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography," *Med Image Anal*, vol. 67, p. 101873, Jan. 2021, doi: 10.1016/J.MEDIA.2020.101873.

[10] "EchoNet Pediatric." Accessed: Jul. 28, 2023. [Online]. Available: https://echonet.github.io/pediatric/

[11] V. Dinh *et al.*, "Cardiac Ultrasound (Echocardiography) Made Easy: Step-By-Step Guide." Accessed: Dec. 05, 2023. [Online]. Available: https://www.pocus101.com/cardiac-ultrasound-echocardiography-made-easy-step-by-step-guide/

[12] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *Adv Neural Inf Process Syst*, vol. 15, pp. 12077–12090, May 2021, Accessed: Jan. 01, 2025. [Online]. Available: https://arxiv.org/abs/2105.15203v3

[13] M. Liao *et al.*, "Left Ventricle Segmentation in Echocardiography with Transformer," *Diagnostics 2023, Vol. 13, Page 2365*, vol. 13, no. 14, p. 2365, Jul. 2023, doi: 10.3390/DIAGNOSTICS13142365.

[14] T. Sourget, S. N. Hasany, F. Mériaudeau, and C. Petitjean, "Can SegFormer be a True Competitor to U-Net for Medical Image Segmentation?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14122 LNCS, pp. 111–118, Jul. 2023, doi: 10.1007/978-3-031-48593-0_8.