Interpretable Echo Analysis Using Self-Supervised Parcels

Sylwia Majchrowska[1,2], Anders Hildeman[3], Ricardo Mokhtari[4], Philip Teare[5]

[1] Data Science Skills & Partnership, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden
[2] AI Sweden, Gothenburg, Sweden
[3] Data Science & Advanced Analytics, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden
[4] Centre for AI, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, Barcelona, Spain
[5] Centre for AI, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Abstract

This study utilizes Sequence-based Transformer for Robust And Parcelised Training Of echo Recordings (STRAPTOR), an AI framework, employing self-supervised learning for automated semantic segmentation of echocardiogram sequences. Utilizing a Vision Transformer (ViT) trained within the self-DIstillation with NO labels (DINO) model for feature extraction, and a specialized Robust And Parcelised Training Of echo Recordings (RAPTOR) head for segmentation refinement, STRAPTOR demonstrates significant advancements in automating echocardiogram analysis. Our approach effectively segments the left ventricle by identifying and aggregating anatomically relevant subregions across cardiac phases. Validation on EchoNet-Dynamic and CAMUS datasets showcases STRAPTOR's robustness and accuracy above 75%. This research highlights the potential of leveraging general datasets for domain-specific tasks, offering a scalable and interpretable solution for cardiac diagnostics.

1. Introduction

Echocardiography plays a crucial role in non-invasive assessment of cardiac function and structure, yet interpreting echocardiographic data is challenging due to variations in image quality, patient anatomy, and subtle cardiac pathologies. Automating echo analysis through traditional supervised learning methods, while effective, requires extensive labeled datasets, making the collection process time-consuming and costly. The recent introduction of *Self-Supervised Learning* (SSL) in machine learning appears to address some of these challenges, including limited labeled data, expensive manual annotations, and the need for models that generalize well across diverse medical conditions [1–6].

Our investigation focuses on leveraging SSL on a large unlabeled dataset, using the *self-DIstillation with NO labels* (DINO) [7] and modified *Self-supervised Transformer with Energy-based Graph Optimization* (STEGO) [8] models as a backbone. This facilitates the generation of self-segmented outputs, termed "parcels," which identify distinct anatomical sub-regions of the heart and unify to segment chosen anatomical regions.

Our findings demonstrate the exceptional robustness of these self-learned parcels across diverse patient profiles and cardiac cycle phases. in this study, we conducted a comprehensive evaluation of the SSL approach on publicly available datasets, demonstrating its adaptability to a wide range of data requirements and different training strategies.

2. Methodology

In this study, we used *Sequence-based Transformer for Robust And Parcelised Training Of echo Recordings* (STRAPTOR), an AI framework, to perform semantic segmentation of echocardiogram sequences [9]. The framework consists of two main components: a backbone for feature extraction and a downstream model for specific task execution, as shown in Fig. 1.

The backbone is a *Visual Transformer* (ViT) model trained in a self-supervised manner using the DINO [7] algorithm. This approach allows the model to learn useful representations from images without the need for labeled data, relying instead on a self-distillation process where a student model learns to predict the output of a teacher model. In that way, this approach is highly versatile and adaptable to various computer vision tasks. Building upon the semi-latent features provided by the DINO-trained ViT backbone, we applied a modified STEGO head, called *Robust And Parcelised Training Of echo Recordings* (RAPTOR), tailored for echocardiogram sequences anal-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.302

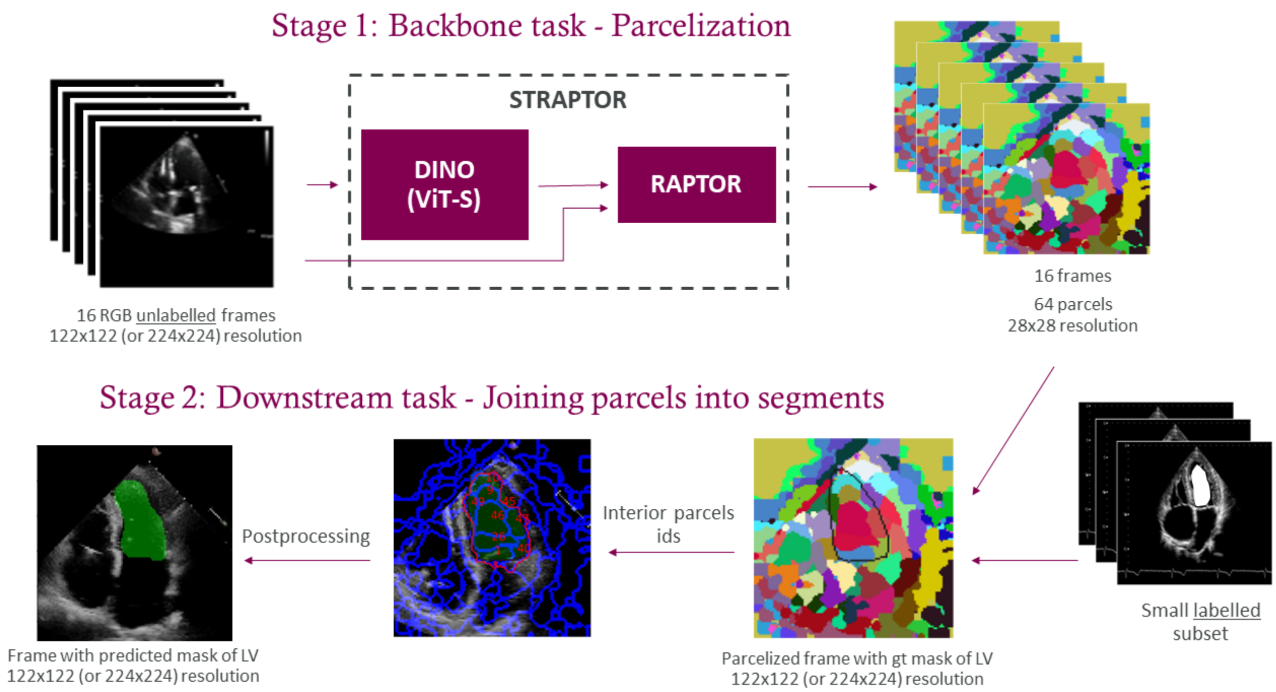

Figure 1. The framework's pipeline comprises two stages: the initial backbone task generates parcelized frames from echocardiographic videos, and the subsequent downstream task combines these parcels to achieve meaningful segmentation of cardiac regions.

ysis, to achieve fully self-supervised image semantic segmentation. The combination of DINO and RAPTOR forms a cohesive framework called STRAPTOR, which produces parcelized frames of echocardiograms.

To refine the segmentation, we unite parcels encompassing the target cardiac subregion, namely *Left Ventricle* (LV), ensuring a single connected component – no holes – and a smooth outline. This is achieved through post-processing steps, including selecting the largest connected segment of interior parcels and refining segmentation edges using an active contour model.

3. Experiments

3.1. Datasets

EchoNet-Dynamic dataset [10] is a publicly available dataset specifically designed for research in echocardiography, a key imaging technique used in the diagnosis and management of heart diseases. This dataset contains 10,030 labeled echocardiogram videos of *Apical 4-Chamber* (A4C) cardiac view. These videos are annotated by clinical experts with mask of LV for *End-Diastole* (ED) and *End-Systole* (ES) frames. The dataset represents a wide range of patients, including variations in age, gender, and heart disease conditions.

Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) dataset [11] consists of echocardiographic videos from 500 patients and includes A4C and *Apical 2-Chamber* (A2C) views of the heart. The sequences were manually labelled with contours of LV and *Left Atrium* (LA) for ED and ES frames.

ImageNet dataset [12] is one of the most influential datasets in the field of computer vision. It was designed for use in visual object recognition software research. ImageNet contains more than 14 million images, categorized into over 20,000 categories. Images in ImageNet are manually annotated with labels and bounding boxes to indicate the presence and position of objects. ImageNet has been pivotal in the development of deep learning models, especially *Convolutional Neural Network* (CNN). **ImageNet-Tiny** [13] is a scaled-down version of the ImageNet dataset, designed for use in scenarios where the computational resources or storage capacities are limited. It contains 120,000 images, with 500 images per class for 200 classes.

Table 1. Reported DSC and IoU as average ± standard deviation for ES and ED frames calculated across the test subsets of EchoNet-Dynamic (n=1277) and CAMUS (n=100). The best results are highlighted in bold.

Training data		**EchoNet-Dynamic [10]**				**CAMUS [11]**			
DINO	**RAPTOR**	$DSC_{LV_{ES}}[\%]$	$DSC_{LV_{ED}}[\%]$	$IoU_{LV_{ES}}[\%]$	$IoU_{LV_{ED}}[\%]$	$DSC_{LV_{ES}}[\%]$	$DSC_{LV_{ED}}[\%]$	$IoU_{LV_{ES}}[\%]$	$IoU_{LV_{ED}}[\%]$
EchoNet-Dynamic [10]	EchoNet-Dynamic [10]	$\mathbf{75.62 \pm 9.36}$	76.94 ± 6.46	$\mathbf{61.61 \pm 10.93}$	62.95 ± 8.26	$\mathbf{68.61 \pm 11.14}$	72.72 ± 8.82	$\mathbf{53.18 \pm 11.60}$	57.87 ± 10.53
EchoNet-Dynamic [10]	ImageNet-tiny [13]	70.75 ± 17.56	73.53 ± 14.62	57.08 ± 17.55	59.97 ± 16.05	60.17 ± 28.92	68.99 ± 16.94	48.10 ± 24.75	54.88 ± 17.47
ImageNet [12]	EchoNet-Dynamic [10]	73.48 ± 10.81	$\mathbf{77.66 \pm 9.16}$	59.09 ± 12.06	$\mathbf{64.32 \pm 11.38}$	66.92 ± 12.22	$\mathbf{72.94 \pm 9.47}$	51.44 ± 12.73	$\mathbf{58.27 \pm 11.51}$
ImageNet [12]	ImageNet-tiny [13]	46.15 ± 27.63	46.13 ± 26.28	33.94 ± 22.23	33.60 ± 21.48	17.95 ± 20.20	53.23 ± 13.14	11.36 ± 13.61	37.32 ± 11.72

3.2. Evaluation metric

Evaluation metrics for segmentation, particularly in the biomedical imaging domain, are crucial for assessing the performance of segmentation algorithms. These metrics help in quantifying the accuracy of the segmentation in comparison to a ground truth (usually manually annotated by experts).

Dice Similarity Coefficient (DSC) is commonly used evaluation metrics in medical image segmentation. It measures the overlap between the predicted segmentation and the ground truth. It is defined as twice the area of overlap divided by the total number of pixels in both the predicted and true segmentation masks. It can by calculated using the formula $\mathrm{DSC} = \frac{2\times|\mathrm{X}\cap\mathrm{Y}|}{|\mathrm{X}|+|\mathrm{Y}|}$, where X is the predicted segmentation and Y is the ground truth. A DSC of 1 indicates perfect overlap, while a DSC of 0 indicates no overlap.

The Jaccard Index, also known as *Intersection over Union* (IoU), is metric similar to DSC, but it measures the intersection over the union of the predicted and true segmentation masks. The formula to calculate IoU states as follow $\mathrm{IoU} = \frac{|\mathrm{X}\cap\mathrm{Y}|}{|\mathrm{X}\cup\mathrm{Y}|}$. The metrics values range from 0 to 1, with 1 being a perfect match and 0 indicating no overlap.

3.3. Ablation Studies

We conducted comprehensive ablation studies to evaluate the effectiveness and necessity of various components within our echocardiographic image segmentation pipeline, which integrates the DINO and RAPTOR models. These studies were designed to assess the impact of different training strategies and dataset choices on the performance of the segmentation task. A key aspect of our ablation studies was to determine the necessity of training any STRAPTOR components specifically on echocardiogram images. We compared the performance of models pretrained on ImageNet with a model pretrained on the EchoNet-Dynamic dataset, adjusting the image and patch sizes to match the computational constraints and resolution requirements.

Our findings revealed that while pretraining on a large, general dataset like ImageNet captures essential latent features, fine-tuning or pretraining on domain-specific datasets significantly enhances the model's sensitivity and performance on echocardiographic image segmentation tasks. This underscores the importance of domain-specific training for the model's head in capturing nuanced features relevant to echocardiography.

Furthermore, chosen evaluation metrics provided insights into the segmentation accuracy. The IoU score was found to be approximately few percentage points lower than the DSC, reflecting the IoU's higher sensitivity to discrepancies in boundary delineation. This difference highlights the complementary nature of these metrics in evaluating segmentation performance, emphasizing the need for a multi-metric approach to fully understand model strengths and limitations [14].

Additionally, our results on the CAMUS dataset were not as robust as those on the EchoNet-Dynamic dataset. This discrepancy can be attributed to the model being trained solely on EchoNet-Dynamic data, which comprises only apical four-chamber views and possesses different image quality characteristics compared to the much smaller CAMUS dataset, which includes both A4C and A2C views. This observation further reinforces the critical role of in-domain training data and the necessity of considering dataset-specific characteristics in model training and evaluation strategies.

4. Conclusion

In this study, we have demonstrated the potential of SSL for the analysis of echocardiograms by generating interpretable parcels. By integrating a ViT trained via the DINO algorithm for robust feature extraction and a RAPTOR head for semantic segmentation, we have demonstrated a step forward in the automation of echocardiogram analysis. Our findings underscore the effectiveness of STRAPTOR in segmenting the left ventricle, showcasing its ability to accurately identify and aggregate anatomically relevant subregions across various cardiac phases. The validation of our framework on the EchoNet-Dynamic and CAMUS datasets confirms its adaptability and robustness.

Moreover, this research illuminates the potential of leveraging pre-trained models on general datasets, such as ImageNet, for domain-specific tasks, thereby reducing the need for extensive labeled datasets in medical imaging. The success of STRAPTOR in echocardiogram segmentation exemplifies the transformative impact of self-supervised learning in medical image analysis, paving the way for more scalable, accurate, and interpretable diagnostic tools in cardiac healthcare.

As we look to the future, the STRAPTOR framework opens new avenues for research and application in the field of cardiac imaging. Its adaptability to different datasets and cardiac structures suggests a wide range of potential uses beyond the left ventricle segmentation, offering a promising direction for further exploration and development. Ultimately, the advancements made through this study contribute significantly to the ongoing efforts to enhance the precision and efficiency of cardiac diagnostics, with the potential to improve patient outcomes and streamline clinical workflows.

Acknowledgments

This work has been carried out during the *Eye for AI Program* thanks to the support of AstraZeneca and AI Sweden. Currently S.M. is a fellow of the AstraZeneca Postdoctoral Research Programme.

References

[1] Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D. Self-supervised learning for medical image analysis using image context restoration. Medical Image Analysis 2019;58:101539. ISSN 1361-8415.

[2] Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 2021;43(11):4037–4058.

[3] Pérez-García F, Dorent R, Rizzi M, Cardinale F, Frazzini V, Navarro V, Essert C, Ollivier I, Vercauteren T, Sparks R, Duncan JS, Ourselin S. A self-supervised learning strategy for postoperative brain cavity segmentation simulating resections. Int J Comput Assist Radiol Surg 2021; 16(10):1653–1661.

[4] Ding X, Liu Z, Li X. Free lunch for surgical video understanding by distilling self-supervisions. In Wang L, Dou Q, Fletcher PT, Speidel S, Li S (eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Cham: Springer Nature Switzerland, 2022; 365–375.

[5] Saeed M, Muhtaseb R, Yaqub M. Contrastive pretraining for echocardiography segmentation with limited data. In Yang G, Aviles-Rivero A, Roberts M, Schönlieb CB (eds.), MIUA. Cham: Springer. ISBN 978-3-031-12053-4, 2022; 680–691.

[6] Holste G, Oikonomou EK, Mortazavi BJ, Wang Z, Khera R. Efficient deep learning-based automated diagnosis from echocardiography with contrastive self-supervised learning. Commun Med Lond 2024;4(1):133.

[7] Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, Joulin A. Emerging properties in self-supervised vision transformers. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 10 2021; 9630–9640.

[8] Hamilton M, Zhang Z, Hariharan B, Snavely N, Freeman WT. Unsupervised semantic segmentation by distilling feature correspondences. In The Tenth International Conference on Learning Representations, (ICLR). OpenReview.net, 2022; .

[9] Majchrowska S, Hildeman A, Teare P, Diethe T. Unlocking the heart using adaptive locked agnostic networks. CoRR 2023;abs/2309.11899.

[10] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Curt P. Langlotz PAH, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based ai for beat-to-beat assessment of cardiac function. Nature 03 2020;580.

[11] Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, Espeland T, Berg E, Jodoin PM, Grenier T, Lartizien C, D'hooge J, Løvstakken L, Bernard O. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging 08 2019;38(9):2198–2210.

[12] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009; 248–255.

[13] Le Y, Yang XS. Tiny imagenet visual recognition challenge. 2015; URL `https://api.semanticscholar.org/CorpusID:16664790`.

[14] Hu H, Shen L, Guan Q, Li X, Zhou Q, Ruan S. Deep co-supervision and attention fusion strategy for automatic covid-19 lung infection segmentation on ct images. Pattern Recognition 2022;124:108452. ISSN 0031-3203.

Address for correspondence:

Sylwia Majchrowska
Pepparedsleden 1, 431 83 Mölndal
sylwia.majchrowska1@astrazeneca.com