IRSTI 50.41.25

10.51889/2959-5894.2024.85.1.014

B.M. Ukibassov[1], S.B. Rakhmetulayeva[1]*, A.K. Bolshibayeva[1]
[1] International Information Technology University, Almaty, Kazakhstan
*e-mail: ssrakhmetulayeva@gmail.com

SEMANTIC SEGMENTATION DEEP LEARNING MODELS IN ECHOCARDIOGRAPHY: CUSTOM DATASET-BASED FINE-TUNING

Abstract

This study enhances the accuracy of semantic segmentation models in cardiology using a custom echocardiogram dataset. The goal is to adapt an existing deep learning model for better segmentation of heart structures in echocardiographic images, crucial for automated cardiac disease diagnosis. The performance improvement is evaluated using cardiology-specific metrics, showing enhanced segmentation accuracy of cardiac structures. This approach increases the model's clinical utility for cardiologists in diagnostics and treatment planning. The results highlight the potential of customized deep learning models in medical imaging and emphasize the importance of specialized datasets for precision in medical applications. This research contributes significantly to artificial intelligence in healthcare, offering advancements in automated echocardiographic analysis for clinical use.

Keywords: self-supervised clustering network, segmentation, synthetic datasets, echocardiogram image processing.

Б.М. Укибасов[1], С.Б. Рахметулаева[1], А.К. Болшибаева[1]
[1] Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

ЭХОКАРДИОГРАФИЯДАҒЫ СЕМАНТИКАЛЫҚ СЕГМЕНТТЕУДІ ТЕРЕҢ ОҚЫТУ МОДЕЛЬДЕРІ: ДЕРЕКТЕР ЖИЫНТЫҒЫНА НЕГІЗДЕЛГЕН FINE-TUNING

Аңдатпа

Бұл мақалада эхокардиограмма деректер жинағын пайдалана отырып, кардиологиядағы семантикалық сегменттеу үлгілерінің дәлдігін жақсартады. Зерттеудің негізгі мақсаты - жүрек ауруларын автоматтандырылған диагностикалау үшін маңызды болып табылатын эхокардиографиялық кескіндердегі жүрек құрылымдарын дәлірек сегментациялау үшін терең оқыту модельдерін бейімдеу. Модельдің жұмысының жақсаруы жүрек құрылымдарының сегментациясының дәлдігін көрсететін кардиологияға тән көрсеткіштерді қолдану арқылы бағаланады. Бұл тәсіл диагностика мен емдеуді жоспарлауда кардиологтар үшін модельдің клиникалық пайдалылығын арттырады. Алынған нәтижелер негізігде медициналық бейнелеудегі терең оқыту үлгілерінің әлеуетін және медициналық қолданбалардағы дәлдікті қамтамасыз ету үшін арнайы деректер жиынының маңыздылығын көрсетеді. Бұл зерттеу клиникалық қолдану үшін автоматтандырылған эхокардиографиялық талдаудың жетістіктерін ұсына отырып, денсаулық сақтау саласындағы жасанды интеллекттің дамуына айтарлықтай үлес қосады.

Түйін сөздер: өзін-өзі реттейтін кластерлік желі, сегменттеу, синтетикалық деректер жиынтығы, эхокардиограмма кескінін өңдеу.

Б.М. Укибасов[1], С.Б. Рахметулаева[1], А.К. Болшибаева[1]
[1] Международный университет информационных технологий, г.Алматы, Казахстан

МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ В ЭХОКАРДИОГРАФИИ: FINE-TUNING НА ОСНОВЕ НАБОРА ДАННЫХ

Аннотация

Данная статья посвящена эксперименту по повышению точности моделей семантической сегментации в кардиологии с использованием специального набора данных эхокардиограммы. Цель состоит в том, чтобы адаптировать существующую модель глубокого обучения для лучшей сегментации структур сердца на эхокардиографических изображениях, что имеет решающее значение для автоматизированной диагностики заболеваний сердца. Улучшение производительности оценивается с использованием показателей, специфичных для кардиологии, которые показывают повышенную точность сегментации сердечных структур. Такой подход повышает клиническую полезность модели для кардиологов при диагностике и планировании лечения. Результаты подчеркивают потенциал индивидуальных моделей глубокого обучения в медицинской визуализации и подчеркивают важность специализированных наборов данных для точности в медицинских приложениях. Это исследование вносит значительный вклад в развитие искусственного интеллекта в здравоохранении, предлагая достижения в области автоматизированного эхокардиографического анализа для клинического использования.

Ключевые слова: самоконтролируемая кластерная сеть, сегментация, синтетические наборы данных, обработка изображений эхокардиограммы.

Introduction

Semantic segmentation involves the classification of every individual pixel inside a picture. It may be seen as a more accurate method of categorising a picture. It offers a broad spectrum of applications in areas such as medical imaging and autonomous driving. For instance, in the case of our pizza delivery robot, it is crucial to accurately determine the precise location of the sidewalk inside a picture, rather than just determining its presence or absence.

Given that semantic segmentation falls under the category of classification, the network designs used for image classification and semantic segmentation exhibit significant similarities. In 2014 [1] published a groundbreaking article that used convolutional neural networks to do semantic segmentation. In recent times, Transformers have found use in image classification, such as the ViT model. Furthermore, they are currently being employed for semantic segmentation, therefore advancing the current state-of-the-art.

SegFormer is a semantic segmentation model proposed by [2] in 2021. The model has a hierarchical Transformer encoder that does not use positional encodings, distinguishing it from ViT. Additionally, it incorporates a straightforward multi-layer perceptron decoder. SegFormer demonstrates exceptional performance on many widely-used datasets.

The article [3] makes a substantial contribution to the field by extending prior research and highlighting the capabilities of sophisticated deep learning models in achieving precise and effective left ventricle analysis. Nevertheless, providing a more in-depth exploration of the clinical implications, addressing uncertainties related to additional features, assessing generalizability in complex cases, and comparing the findings to recent advancements would not only enhance the paper's impact but also offer valuable guidance for future research in this promising domain.

Several recent studies have utilised convolutional neural networks (convnets) to address dense prediction problems. These include semantic segmentation, as demonstrated by [4-6]; boundary prediction for electron microscopy, as shown by [2]; boundary prediction for natural images, as demonstrated by a hybrid model combining a neural network and nearest neighbour approach and image restoration and depth estimation, as demonstrated by [5, 6].

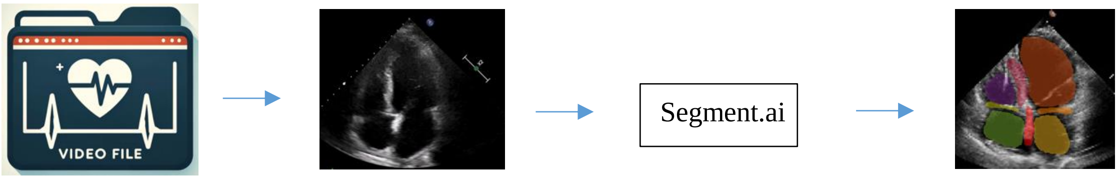
Research methodology

Fine-Tune a semantic segmentation is a crucial technique in the field of computer vision and image processing, where the goal is to understand and label each pixel in an image. Unlike image classification, where the entire image is assigned a single label, or object detection, which identifies and locates objects within an image, semantic segmentation goes a step further by classifying each pixel into a predefined category. This leads to a much finer, pixel-level understanding of the image [7].

Semantic segmentation, a critical technique in computer vision and image processing, is renowned for its high precision and pixel-level accuracy. This precision is essential for accurately understanding the shape and boundaries of objects within an image, making semantic segmentation indispensable in various applications.

In the healthcare sector, semantic segmentation plays a pivotal role, particularly in the analysis of medical scans such as MRIs and CT scans. It aids in the identification and segmentation of different biological structures and anomalies, such as tumors, thereby assisting in diagnosis, treatment planning, and the monitoring of disease progression [5].

A standard 2-dimensional (2D) transthoracic echocardiography acquires several movies from various perspectives using 2D cross-sectional pictures. Video clips may not always include labelled views, so view identification is a crucial process in AI applications. This step is necessary before training the deep neural network and carrying out automated measurement and picture interpretation (Figure 1). Proper selection of the correct perspective is necessary in order to accurately evaluate the echocardiography.Moreover, given that a substantial quantity of annotated data is necessary for training deep learning algorithms, the presence of a view identification algorithm capable of managing this time-consuming and arduous preprocessing stage is anticipated to expedite the progress of artificial intelligence applications in echocardiography.



a) b) c) d)

Figure 1. Data pre-processing and labeling (a - echocardiographic video files; b - .png files; c - data labeling tool; d - labeling process)

Data pre-processing.

Steps:

1. Echocardiographic video files (in .avi format) see Figure 2.
2. Fragmenting echo video file into separate .png files (use for labeling and further model training process).

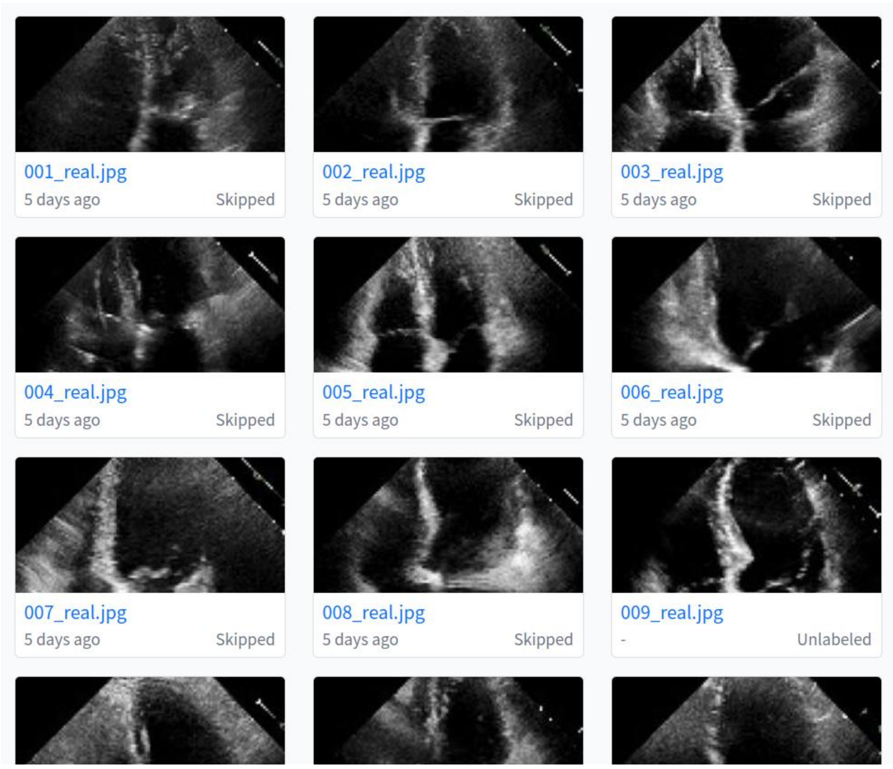


Figure 2. Sample of the fragmented .png files ready for Labeling process

Data labeling for segmentation task.

Steps:

https://segments.ai/ - data labeling tool. Uploading images from previous step into segments.ai tool for further data labeling process.

By the end of the labeling process, we need to release already labeled data in order to use it for the model training process. Segments.ai provides an easy to use interface to release new datasets.

First, we imported all the necessary packages, such as transformers datasets segments-ai evaluate and ect (Figure 3).

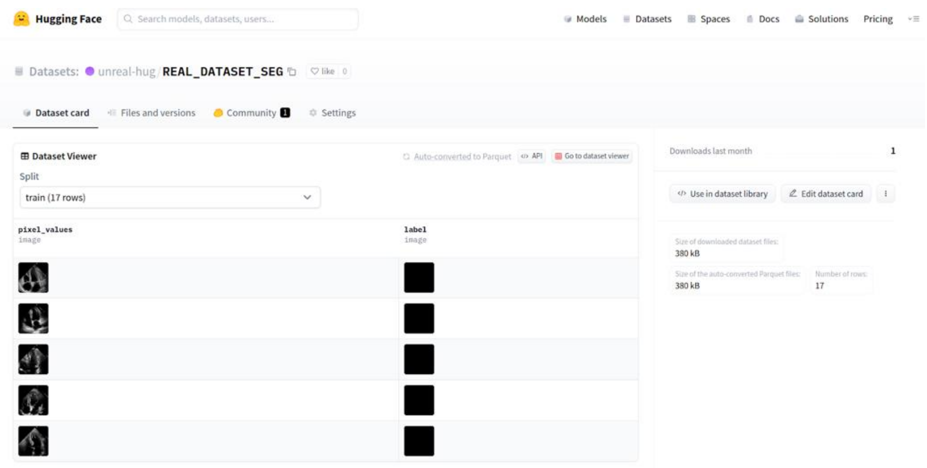


Figure 3. Released dataset on HuggingFace Datasets

Then, we read data from Segments.ai Client as following (Figure 4).

```
from segments import SegmentsClient
from getpass import getpass

api_key = getpass('Enter your API key: ')
segments_client = SegmentsClient(api_key)

Enter your API key: ··········

from segments.huggingface import release2dataset

# dataset_identifier = 'label_h/Test_seg'
dataset_identifier = 'label_h/Real_ECHO_Dateset'
release_name = 'Real_v0.1' #'Test_release_v0.1'

release = segments_client.get_release(dataset_identifier, release_name)
hf_dataset = release2dataset(release)

Map:   0%|          | 0/17 [00:00<?, ? examples/s]
Map:   0%|          | 0/17 [00:00<?, ? examples/s]
```

Figure 4. Block of code how to read data from Segments.ai

Then, we set Training arguments such as defining model type, learning rate, number of epochs, batch size, etc. Finally, we call on our Trainer object function train(). Value can be seen in the following figure (Figure 5). The anatomical structures of the heart are crucial for the process of segmentation. The doctor responsible for maternal-fetal care should manually delineate accurate borders around the heart pictures using the data annotation tool (Segment.ai). Segment.ai is an internet-based annotation tool for constructing picture datasets used in computer vision research. A database of ground truths was created using the notable differences in picture quality, forms, sizes, and orientations among the pregnant women. In foetal echocardiography with normal anatomy, each standard view has a distinct heart chamber structure. Consequently, annotations should be performed for all standard views, including their unique chambers.

```python
from transformers import TrainingArguments

epochs = 50
lr = 0.00006
batch_size = 2

hub_model_id = "segformer-b0-finetuned-segments-ECHO-dev-05-v1"

training_args = TrainingArguments(
    "segformer-b0-finetuned-segments-ECHO-outputs",
    learning_rate=lr,
    num_train_epochs=epochs,
    per_device_train_batch_size=batch_size,
    per_device_eval_batch_size=batch_size,
    save_total_limit=3,
    evaluation_strategy="steps",
    save_strategy="steps",
    save_steps=20,
    eval_steps=20,
    logging_steps=1,
    eval_accumulation_steps=5,
    load_best_model_at_end=True,
    push_to_hub=True,
    hub_model_id=hub_model_id,

    hub_strategy="end",
)
```

Figure 5. Training objects

Examples of such chambers include: The 4CH standard view includes the following heart chambers: aortic (Ao), left atrium (LA), left ventricle (LV), right atrium (RA), and right ventricle (RV), autism spectrum disorder (ASD), ventricular septal defect (VSD), tricuspid valve (TK), mitral valve (MK) and aortic stenosis (AS) (Figure 6).

ReadMe Edit

Real_ECHO_Dateset

Labeling specification Edit

id	name	description	has instances	attributes
1	LV	Left ventricle	yes	yes
2	RV	Right ventricle	yes	yes
3	RA	Right atrium	yes	yes
4	LA	Left atrium	yes	yes
5	ASD	Autism spectrum disorder	yes	yes
6	VSD	Ventricular septal defect	yes	yes
7	TK	Tricuspid valve	yes	yes
9	MK	Mitral valve	yes	yes
10	AS	Aortic stenosis	yes	yes

Figure 6. Labeling specification

Results of the study

Specifically for the identification of heart defects, only the four-chamber (4CH) view was used to analyse pictures of atrial septal defects (ASD), ventricular septal defects (VSD), and atrioventricular septal defects (AVSD). Annotated pictures provide a visual representation of the location of a defect in either the atrium, ventricle, or both. Figure 5 illustrates a collection of annotated photos representing the conventional views of 4CH. Figure 5 displays annotated images highlighting the positions of defects in ASD, VSD, and AVSD. Ultimately, the whole set of 401 tagged photos is designated as the ground truth database and thereafter stored with next statistical information (Figure 7).

Figure 7. Label statistics

Expert professionals were selected to take part in the image labelling procedure: a top-tier doctor from the Mediterra private medical centre with expertise in cardiac surgery, and a distinguished doctor specialising in cardiology from the Research Institute of Cardiac Surgery. This guaranteeed a superior standard of professionalism and precision in the annotation of medical images, which is crucial for attaining optimal outcomes in the procedure.

Discussion

Echocardiography requires not just imaging expertise, but also depends heavily on subjective interpretation. The potential of AI in echocardiographic interpretation lies in its ability to retrieve non-apparent information, thereby offering promising prospects. An instance of this is the subjective and experience-dependent nature of early identification of congenital heart abnormalities by the visual interpretation of right and left ventricular excursion. An objective classification approach is anticipated to enhance the identification of congenital cardiac abnormalities in clinical settings. A recent research demonstrated that a DL-based algorithm for diagnosing congenital deformity achieved a diagnostic accuracy that was similar to the evaluation made by experts. AI-based echocardiographic evaluation may not be vital for proficient individuals. Nevertheless, the attractiveness of having an objective and quantitative evaluation that eliminates errors caused by different observers is significant, as it may greatly improve the efficiency of everyday clinical treatment. Enhancements in these models would enable their use not only in resting-state echocardiography but also in stress echocardiography, hence broadening their therapeutic applicability.

Conclusion

This article contributes a substantial progress in the area of cardiac imaging by improving the precision of semantic segmentation models via the use of a carefully selected echocardiography dataset. Through the process of modifying and optimising an already established deep learning model, we have successfully attained exceptional segmentation of heart structures. This accomplishment represents a crucial advancement in the direction of completely automated detection of cardiac diseases. The use of cardiology-specific measures for performance assessment highlights the significant improvement in the precision of heart structure segmentation.

Importantly, the improved method enhances the practicality of the model, offering cardiologists a dependable instrument for diagnosing and arranging treatments. The results of this study highlight the significant capabilities of personalised deep learning models in the field of medical imaging, with a specific focus on the importance of tailored datasets to ensure accuracy in medical applications.

This study represents a significant advancement in the incorporation of artificial intelligence into the field of healthcare. This enables significant improvements in automated echocardiographic analysis, hence assisting doctors in providing timely and precise cardiac treatment. The findings of this study have broader significance in the field of cardiology, serving as a catalyst for future research to use comparable methodology in other medical imaging techniques. This has the potential to significantly transform diagnostic processes in several areas of healthcare.

Acknowledgements

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP13068032 - Development of methods and algorithms for machine learning for predicting pathologies of the cardiovascular system based on echocardiography and electrocardiography).

References

[1] Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. Computer Vision and Pattern Recognition (cs.CV), https://doi.org/10.48550/arXiv.1411.4038

[2] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Ping Luo. Computer Vision and Pattern Recognition (cs.CV); Machine Learning (cs.LG), https://doi.org/10.48550/arXiv.2105.15203

[3] G.Carneiro, J. C. Nascimento, and A. Freitas, "The Segmentation of the Left Ventricle of the Heart From Ultrasound Data Using Deep Learning Architectures and Derivative-Based Search Methods," IEEE Transactions on Image Processing, vol. 21, no. 3, pp. 968–982, mar 2012.

[4] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2018.

[5] S.B. Rakhmetulayeva., A.K. Bolshibayeva., A.K. Mukasheva, B. Ukubassov, Z.O. Zhanabekov, Diaz, D. Machine learning methods and algorithms for predicting congenital heart pathologies. 17th IEEE International Conference on Application of Information and Communication Technologies, AICT 2023 - Proceedings, 2023

[6] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano. Toward automatic phenotyping of developing embryos from videos. Image Processing, IEEE Transactions on, 14(9):1360–1371, 2019.

[7] P.H. Pinheiro and R. Collobert. Recurrent convolutional neural networks for scene labeling. In ICML, 2014.

[8] Mukasheva, A., Koishiyeva, D., Suimenbayeva, Z., Rakhmetulayeva, S., Bolshibayeva, A., Sadikova, G.(2023). Com- parison evaluation of unet-based models with noise augmentation for breast cancer segmentation on ultrasound images. Eastern-Eu- ropean Journal of Enterprise Technologies, 5 (9 (125)), 85–97. doi: https://doi.org/10.15587/1729-4061.2023. 289044