A Multi-Stage Framework for Simultaneous Digitization and Classification of Electrocardiogram Images

Jingsu Kang[1], Hao Wen[2]

[1]Tianjin Medical University, Tianjin, China
[2]College of Science, China Agricultural University, Beijing, China

Abstract

Aim: This paper addresses the challenge of simultaneous digitization and arrhythmia classification of electrocardiograms (ECGs) captured from images, as presented by the George B. Moody PhysioNet Challenge 2024.

Methods: We propose a multi-stage framework to digitize and classify ECG images concurrently. The framework consists of 3 modules: (1) Region of Interest (ROI) extraction, (2) Arrhythmia classification, and (3) ECG waveform digitization. The ROI extraction module uses a DETR model with a ResNet-50 backbone to locate the ECG region. The classification module uses a lightweight ConvNeXt model to classify the ECG signal. The digitization module uses a U-Net model to extract the ECG waveform from the ROI. The digitization task involves a highly unbalanced segmentation problem, where the ECG waveforms occupy a small portion of the image. We propose a novel loss function to address this issue.

Results: Our team ("Revenger") received a score of 0.33 for the classification problem on the hidden data, ranked 9/16; and a signal-to-noise ratio (SNR) of -0.733 for the digitization problem, ranked 13/16.

Conclusion: The proposed framework demonstrates the feasibility of digitizing and classifying ECG images concurrently, but needs further improvement to achieve better performance.

1. Introduction

Although deep learning has gained significant popularity in the field of electrocardiogram (ECG) analysis, the primary focus has been on digital ECG signals. However, paper printouts (images) of ECG signals remain widely used in clinical practice, particularly in developing countries. The direct analysis and digitization of ECG images are crucial for further analysis using deep learning models.

The George B. Moody PhysioNet Challenge 2024 [1, 2] has posed the challenge of developing automated methods for the simultaneous digitization and arrhythmia classification of ECG images, aiming to advance the development of deep learning models for ECG image analysis. In this work, we developed a multi-stage framework to address this challenge.

2. Methods

2.1. Data Processing

The Challenge did not provide any ECG images directly for method development. Instead, it offered a tool named ECG-Image-Kit [3, 4] to generate ECG images from digital ECG signals. We utilized this tool on the PTB-XL dataset [5] and generated a total of 21,799 synthetic ECG images for method development. Additionally, the PTB-XL+ dataset [6] was used to generate classification labels for the ECG images.

The image generation setup included augmentations such as random Gaussian noise (zero mean and standard deviation randomly chosen within 1 - 50), random color temperature (2000 - 4000 or 10000 - 20000) adjustments, and distortions including handwritten notes, wrinkles, and creases, all conducted independently randomly with 50 % probability. Rotation augmentation was excluded from the generation setup since the Challenge mainly focuses on "normal" ECG images. Bounding boxes for the printed ECG waveforms, lead names, and plotted pixel coordinates of the ECG waveforms were recorded in metadata files alongside the generated images. These supplementary data were used to:
(1) Train the object detection model to detect the bounding boxes of the ECG waveforms and lead names in the ECG images, thus identifying the region of interest (ROI) for subsequent ECG signal extraction.
(2) Train the segmentation model to digitize the ECG waveforms from the images.

The PTB-XL dataset assigned a "strat_fold" property (with values 1-10) to each recording, which we used to split the dataset into training, and validation sets. The validation set has "strat_fold" values 10, as recommended by the PTB-XL dataset, serving for model selection and hy-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.128

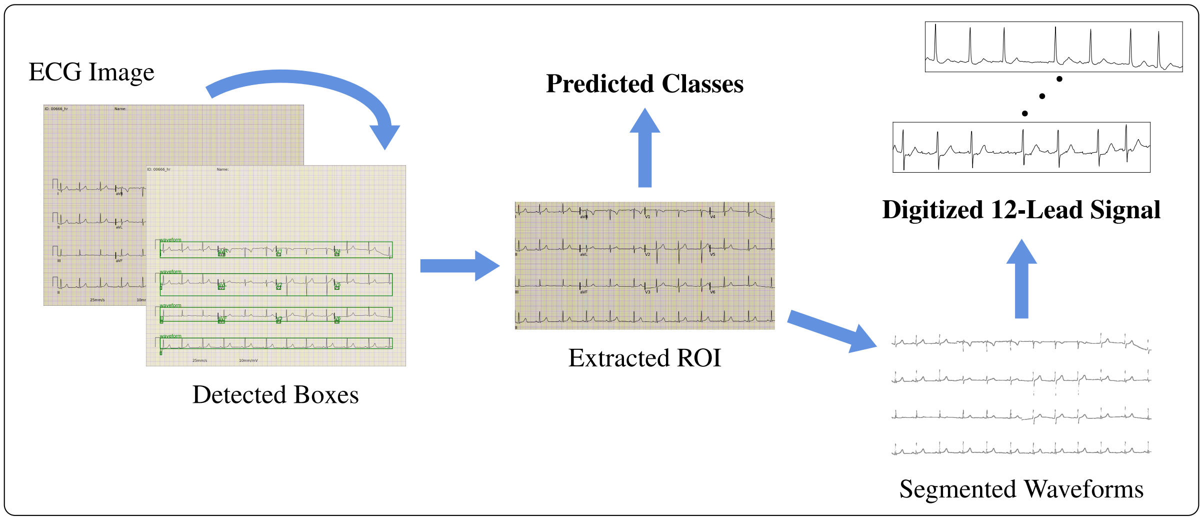

Figure 1: The multi-stage framework for simultaneous digitization and classification of ECG images.

perparameter tuning. The training set comprises the remaining recordings.

2.2. The Multi-Stage Framework

To simultaneously digitize and classify ECG images, we designed a multi-stage framework, as shown in Figure 1. The framework comprises three deep learning-based modules: object detection, segmentation, and classification.

The object detection module identifies the bounding boxes of the ECG waveforms and lead names in the ECG images, thereby determining the region of interest (ROI) for subsequent ECG signal extraction and classification. This module constitutes the first stage of the framework.

The segmentation module extracts a mask of the ECG waveforms from the ROI, which is then transformed into digitized values based on the pixel coordinates of the predicted mask. The bounding boxes obtained from the object detection module are also used to separate the ECG waveforms at different horizontal positions. In cases where the bounding boxes overlap, which commonly occurs when the ECG signals have large amplitudes, we employ a 4-cluster K-means algorithm on the y-coordinates of the predicted mask to locate the isoelectric line of the ECG waveforms and then perform the separation. Concurrently, the classification module predicts the classes of the ECG waveforms from the ROI. Together, these two modules constitute the second stage of the framework.

2.3. Model Selection

We present our criteria for selecting the deep learning models for the three modules discussed in Section 2.2. The models are chosen based on their performance, efficiency, and compatibility with ECG image data. After evaluating several models, we choose the following models for the three modules:

(1) Object Detection: A DETR model [7] with a ResNet-50 [8] backbone.
(2) Classification: A "nano" size ConvNeXt model [9].
(3) Segmentation: A U-Net model [10].

The object detection and classification models are trained from pre-trained models loaded from Huggingface's model hub [11], while the segmentation model is trained from scratch. The models are implemented and trained using the "torch_ecg" library [12]. The sizes in terms of the number of trainable parameters of the models are listed in Table 1.

	ConvNeXt-nano	DETR-ResNet-50	U-Net
# Params	15M	41M	31M

Table 1: Number of trainable parameters in the models.

For classification, using an oversized model may not improve the performance and can even degrade it due to overfitting. Therefore, we opted for a lightweight model to balance performance and efficiency. The evaluation metrics (in terms of F1 score) of different sizes of ConvNeXt models on the left-out validation set, along with their number of trainable parameters, are presented in Table 2.

The DETR model achieved a mean average precision (mAP) of 0.90 and the U-Net model achieved a Dice coefficient of 0.73, both on the left-out validation set.

2.4. Loss Functions

Given the highly imbalanced data distribution for the classification problem, as illustrated in Figure 2, and the

Threshold	size of the ConvNeXt model			
	nano	base	large	nano (head only)
0.5	0.287	0.243	0.208	0.216
0.7	0.344	0.284	0.207	0.213
0.9	**0.428**	0.382	0.205	0.237
0.95	**0.450**	**0.448**	0.197	0.218
# Params	15M	88M	197M	0.3M

Table 2: Evaluation metrics of ConvNeXt models of different sizes. The values are F1 scores on the left-out validation set. The threshold refers to the probability threshold for the multi-label predictions. "head only" refers to the model with backbone weights frozen. The last row shows the number of trainable parameters in each model.

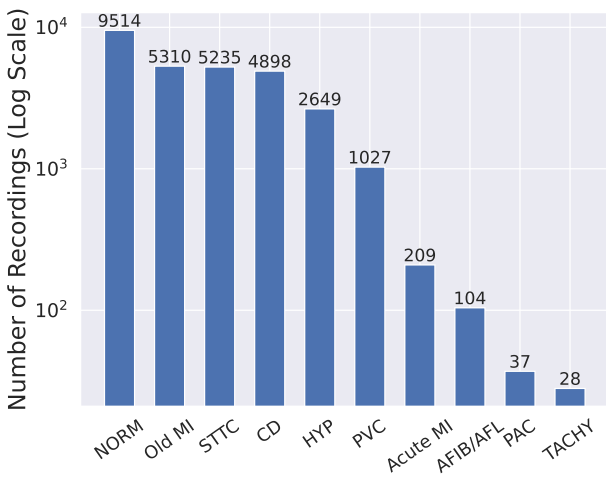

Figure 2: Data distribution of the PTB-XL dataset. Labels are generated from the PTB-XL+ dataset. Abbreviations: NORM, Normal; OLD MI, Old Myocardial Infarction; STTC, ST/T Changes; CD, Conduction Disturbances; HYP, Hypertrophy; PVC, Premature Ventricular Complex; Acute MI, Acute Myocardial Infarction; AFIB/AFL, Atrial Fibrillation/Atrial Flutter; PAC, Premature Atrial Complex; TACHY, Tachycardia.

multi-label nature of the task, we adopted the asymmetric loss function [13], as defined in Equation (1).

$$\begin{aligned} L = & -y \cdot (1-p)^{\gamma_+} \log(p) \\ & - (1-y) \cdot (p_m)^{\gamma_-} \log(1-p_m), \qquad (1) \\ \text{where} \quad & p_m = \max(p-m, 0), \end{aligned}$$

where y is the ground truth label, p is the predicted probability, m is the probability margin, and γ_+ and γ_- are the positive and negative focusing parameters, respectively. The margin m and the focusing parameters γ_+ and γ_- are hyperparameters that can be tuned. We set $\gamma_+ = 1$, $\gamma_- = 4$, and $m = 0.05$ in our experiments.

For the segmentation task, considering that the ECG waveforms occupy only a small portion of the image, we used the binary cross-entropy loss multiplied by a weight mask to balance the loss contribution from different parts of the ECG waveforms. The weight mask is the sum of multiple dilations (for example, dilation iterations 2, 4, 6, and 8) of the binarized ground truth mask. An example of the weight mask is shown in Figure 3.

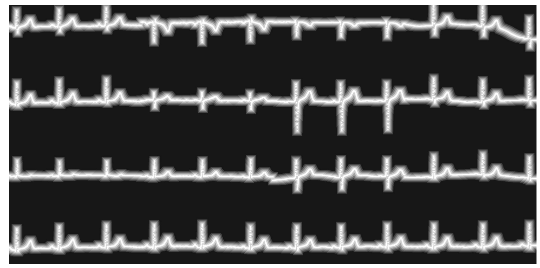

Figure 3: An example of the weight mask for the segmentation loss function.

For the object detection task, we used the standard DETR loss function, which is a linear combination of the cross-entropy loss for the bounding box class predictions (with a coefficient of 1) and the bounding box coordinate regression loss (L1 loss with a coefficient of 5 and generalized IoU loss with a coefficient of 2).

3. Results

Task	Score	Rank
Classification	F-measure: 0.33	9/16
Digitization	SNR: -0.733	13/16

Table 3: F1 score and SNR for our selected entry (team "Revenger") on the classification and digitization tasks, respectively. The rankings are based on the hidden data.

The results of our team ("Revenger") on the hidden data [14] are presented in Table 3 . Results on public training data (specifically, the left-out validation set) are also provided for reference: F-measure: 0.45; SNR: 1.21. Our team's performance is in the middle of all teams.

4. Discussion and Conclusions

The results presented in Section 3 indicate that our proposed method is relatively effective in digitizing and classifying ECG images concurrently, achieving acceptable SNR and F1 scores. Our framework (ref. Fig. 1) offers a lightweight and efficient solution to this challenge. The average inference time of the entire pipeline is approximately 2 seconds per ECG image, measured on a single NVIDIA

RTX 3090 GPU. This capability of processing ECG images in real time is crucial for clinical applications.

However, there are observable gaps between the SNR metric on the public training data and the hidden data. A possible reason for this discrepancy is the difference in data distribution (format). The training data generated using ECG-Image-Kit [3, 4] all follow the standard 3-by-4 format, namely 3 rows and 4 columns of the 12-lead ECG waveforms (each lead having a 2.5s length), with an additional row for a full-length lead II waveform. The post-processing step after segmentation is designed to handle this specific format. However, the hidden data may have different formats, as inferred from the released samples. For example, there could be additional full-length waveforms in the images. This discrepancy may lead to a performance drop on the hidden data. To address this issue, we can consider designing a more flexible post-processing step that can handle various ECG image formats and make use of the detected lead names by the object detection model to determine the format of the ECG image, which is not used in the current framework.

Our method can be further improved in several aspects. First, the object detection module can be enhanced by incorporating a more effective backbone network or a more advanced architecture, including the latest YOLO series models. An end-to-end digitization model from the segmented ECG waveform masks or even directly from the ROI can be developed to simplify the pipeline, replacing the current rule-based conversion from pixel coordinates to digitized values. Finally, handling rotated ECG images can be considered to improve the robustness of the framework, which is beyond the scope of the current work.

Acknowledgments

The authors thank Professor Deren Han from Beihang University for helping accomplish this work.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[2] Reyna MA, Deepanshi, Weigle J, Koscova Z, Elola A, Seyedi S, et al. Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024. Computing in Cardiology 2024;51:1–4.

[3] Shivashankara KK, Deepanshi, Mehri Shervedani A, Clifford GD, Reyna MA, Sameni R. ECG-Image-Kit: A Synthetic Image Generation Toolbox to Facilitate Deep Learning-Based Electrocardiogram Digitization. Physiological Measurement 5 2024;45(5):055019. ISSN 1361-6579.

[4] Deepanshi, Shivashankara KK, Clifford GD, Reyna MA, Sameni R. ECG-Image-Kit: A Toolkit for Synthesis, Analysis, and Digitization of Electrocardiogram Images, 1 2024. Online at: `https://github.com/alphanumericslab/ecg-image-kit`.

[5] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[6] Strodthoff N, Mehari T, Nagel C, Aston PJ, Sundar A, Graff C, et al. PTB-XL+, a comprehensive electrocardiographic feature dataset. Scientific Data 5 2023;10(1). ISSN 2052-4463.

[7] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-End Object Detection with Transformers. In Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-030-58451-1, 2020; 213–229.

[8] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Institute of Electrical and Electronics Engineers (IEEE), 6 2016; 770–778.

[9] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 6 2022; 11966–11976.

[10] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer Assisted Intervention. Springer, 2015; 234–241.

[11] Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Online: Association for Computational Linguistics, 10 2020; 38–45.

[12] Wen H, Kang J. A Novel Deep Learning Package for Electrocardiography Research. Physiological Measurement 11 2022;43(11):115006. ISSN 1361-6579.

[13] Ridnik T, Ben-Baruch E, Zamir N, Noy A, Friedman I, Protter M, et al. Asymmetric Loss for Multi-Label Classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 10 2021; 82–91.

[14] Reyna MA, Deepanshi, Weigle J, Koscova Z, Campbell K, Shivashankara KK, et al. ECG-Image-Database: A Dataset of ECG Images with Real-World Imaging and Scanning Artifacts; A Foundation for Computerized ECG Image Digitization and Analysis. arXiv240916612v1 9 2024;URL `https://arxiv.org/abs/2409.16612`.

Address for correspondence:

Hao Wen
No. 17, Qinghua East Road, Haidian District, Beijing, China
wenh06@cau.edu.cn,wenh06@gmail.com