Segmentation-based Extraction of Key Components from ECG Images: A Framework for Precise Classification and Digitization

Hong-Cheol Yoon[1], Dong-Kyu Kim[1], Hyun-Seok Kim[2], Woo-Young Seo[3], Chang-Hoe Heo[4], Sung-Hoon Kim[1,4]

[1]Department of Anesthesiology and Pain Medicine, Asan Medical Center, Brain Korea 21 Project, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Department of Anesthesiology and Pain Medicine, University of Ulsan College of Medicine, Seoul, Republic of Korea
[3] Biomedical Engineering Research Center, Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea
[4] Signal House Co., Ltd, Seoul, Republic of Korea

Abstract

The physical and paper Electrocardiography (ECG) contain valuable insights into the history and diversity of cardiovascular diseases (CVDs). The development of algorithms that can digitize and classify these images could significantly improve our understanding and treatment of CVDs, particularly in underrepresented and underserved populations. As part of the George B. Moody PhysioNet Challenge 2024, we propose a deep learning approach to digitize and classify ECG images. Our methodology employs a deep learning segmentation model to extract key components, which are then used to train a classification model for the detection of CVDs and to digitize the signal. Our team, BAPORLab, achieved a signal-to-noise ratio of 5.493 placing 2nd in the digitization task. In the classification task, we achieved a macro F-measure of 0.730, ranked 3rd.

1. Introduction

The George B. Moody PhysioNet Challenge [1, 2] presented a distinctive opportunity for research teams from around the globe to develop algorithms with a focus on the digitization and classification of electrocardiography (ECG) images. The automation of analysis could result in a more timely and accurate detection of major cardiovascular diseases (CVDs), particularly in areas where access to advanced diagnostic tools may be limited. Furthermore, the digitization and classification of ECG images could facilitate the broader usage of historical and non-digital ECG data, thereby enhancing its accessibility for clinical use and further research.

One of the key complexities in this field is the variability in ECG image formats, especially when the data is sourced from real-world, non-standardized environments. ECG images can manifest in a wide array of forms due to differences in imaging techniques, which can lead to variations in geometry, pixel intensity, and overall image quality. This diversity poses a significant challenge for the development of robust algorithms capable of processing these images uniformly. Despite this, many existing studies have relied on internal datasets, which do not fully capture the variability present in real-world data. As a result, models trained on these datasets may have difficulty generalizing to more diverse and noisy real-world ECG images.

In this study, we propose a method for extracting key components—waveform, gridline, and text—from ECG images. These components are then utilized for both classification and digitization tasks. By leveraging this approach, we achieved robust and accurate results using generated datasets. This method not only enhances the performance of ECG analysis but also demonstrates the potential for effective data processing even when relying on synthetic data.

2. Methods

In this section, we describe the detailed methodology used in our study. The overall structure of the proposed architecture is depicted in Figure 1, which provides a comprehensive overview of the process.

2.1 Dataset

The PTB-XL database [3, 4] used for training contains 10-second, 12-lead ECGs with 21,799 clinical records from 18,869 patients. Additionally, we used 62,734 ECG recordings from the PhysioNet Challenge 2021 database [5, 6] and 46,729 ECG recordings from the MIMIC-IV database [7] for pretraining the model.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.227

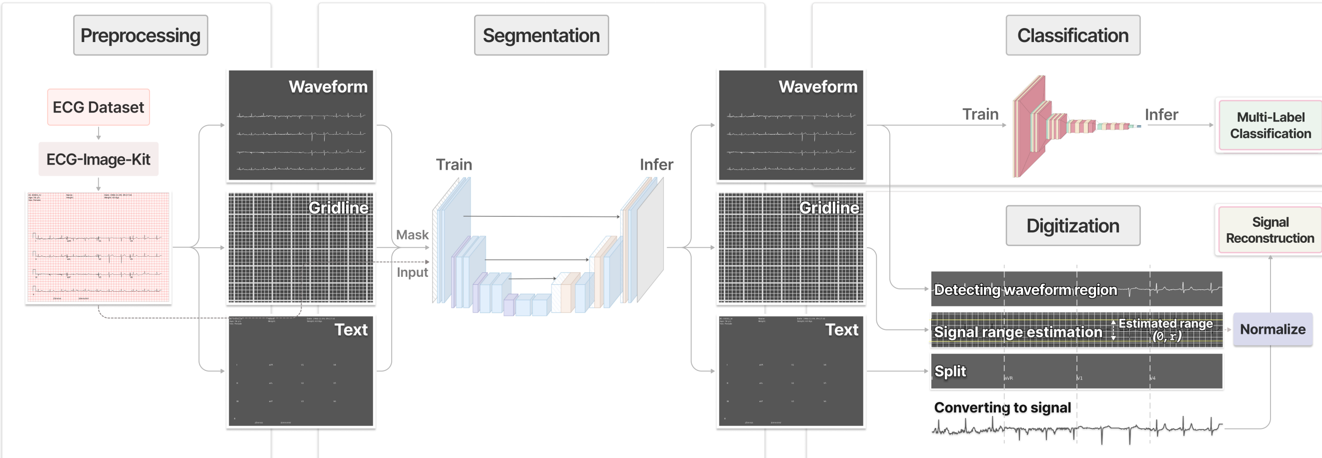

Figure 1. The methodology for extracting key components from generated ECG images using a segmentation model, including the requisite preprocessing steps. These components are employed in the digitization and classification processes.

2.2. Preprocessing

The generation of an undistorted basic ECG image, along with its corresponding annotations, using the ECG-Image-Kit [8, 9] tool is a critical prerequisite for the creation of the masks required for training the segmentation model. In our approach, we found that the basic ECG image generated by the tool could be deconstructed into distinct waveform, gridline, and text channel masks. This separation was achieved through the application of Hue, Saturation, Value transformation, followed by a series of thresholding operations.

Upon closer examination, it was observed that the waveform channel not only contained the primary ECG signal but also included additional elements, such as calibration markers and lead divider markers. It was imperative that these superfluous components be removed to prevent the introduction of noise into the waveform area. To address this challenge, we employed the 'plotted_pixels' data present in the annotations. The data proved invaluable in plotting subsequent waveforms and identifying regions of overlap with the waveform channel. As a result, a more refined and precise waveform mask was generated, free from the interference of non-essential markers. Finally, the masks from the waveform, gridlines, and text channels were combined into a 3-channel.

2.3 Deep learning model

A deep learning-based segmentation model is employed to effectively differentiate between the principal components of ECG images, we utilized pre-processed masks as ground truth during training. The accurate segmentation of components is of great importance, as it enables the model to isolate and extract only the essential information from the ECG data. The extracted information is then employed in subsequent stages of the pipeline, specifically in the classification model and the digitization algorithm.

The segmentation model, which serves as the foundation for the entire workflow, is constructed upon a U-Net based encoder-decoder architectural framework. This design is particularly well-suited for image segmentation tasks, as it enables the model to efficiently extract and capture intricate features through dimensionality reduction, while also allowing for the precise reconstruction of spatial information via up-sampling. The encoder architecture employed, EfficientNet-B0 [10], is responsible for progressively reducing the dimensionality of the input data, thereby distilling the salient features necessary for accurate segmentation. In contrast, the decoder of U-Net is designed to restore the original spatial dimensions, ensuring that the output retains high levels of detail and precision.

In the classification task, we utilize a convolutional neural network (CNN) model with the same architecture as the encoder of the segmentation model. The waveform channel extracted from the mask generated by the segmentation model is fed to the model, ensuring that the features segmented are directly utilized for multi-label classification. Moreover, the same architecture between the classification and segmentation models provides a foundation for future research, such as developing end-to-end models.

2.3.1 Data augmentation

Training with only basic ECG images presents a challenge in reflecting the characteristics of various real-world data sets. To address this issue, we employed a range of data augmentation techniques to better align the training data with the characteristics of the hidden data sets. These techniques were utilized with the objective of enhancing the robustness and generalizability of our models by simulating the variability observed in hidden data sets. To enhance the robustness and generalizability of the segmentation model, we employed a series of techniques, including RGB value shifting, downscaling, and adding noise. Furthermore, we applied a range of geometric transformations, such as cropping, rotating, flipping, perspective warping, and random tiling to both models. These transformations facilitate the model's learning and adaptation to various spatial distortions, enhancing performance across diverse image data.

2.3.2 Experimental settings

We employed a set of hyperparameters and training configurations to optimize the performance of our models. The segmentation model was trained for 12 epochs with an initial learning rate of 3e-4, while the classification model was trained for 16 epochs with an initial learning rate of 5e-4. To ensure the generalization of results, both models were validated through 4-fold cross validation.

Both models were trained using a batch size of 2, which allowed for the effective management of memory constraints while maintaining an optimal learning process. For improving convergence, gradient accumulation performed at 8-steps interval, which stabilized updates by simulating a larger batch size without a notable increase in computational resources.

2.3.3 Pretraining

To further enhance performance, we utilized external datasets for pretraining and applying the same preprocessing methods to ensure consistency. This strategy allowed us to leverage pretrained weights, enabling the segmentation model to achieve its target performance with minimal additional training on the provided dataset, thereby significantly reducing overall training time. Additionally, the classification model exhibited an approximate +0.05 improvement in cross validation score, indicating a substantial boost in performance. This approach not only accelerated the training process but also enhanced the ability to generalize to diverse datasets. The pretraining phase followed the configuration detailed in Section 2.3.2. and the weights with the highest score were fine-tuned using the training data.

2.4 Digitization algorithm

The digitization task is performed applying a rule-based algorithm to the mask generated by the segmentation model. This mask is a three-channel binarized image containing waveform, gridline, and text information. The algorithm is comprised of three principal stages. Subsequently, short and long leads are identified, rearranged and reconstructed to their original signal form. Initially, the waveform region is detected and segmented into long slices. Secondly, the signal range is calculated from the gridline channel of the slices. Thirdly, the waveform channel of slices is converted into a 1-dimensional signal and normalized using the acquired signal range.

Detecting waveform region: Once the waveform region within the extracted mask has been identified, it can be cropped as a slice for efficient post-processing. To achieve this, the waveform channel mask is first skeletonized [11], whereby the pixel thickness is converted to 1. Subsequently, a kernel of size (*height, 1*) is used to perform a left-to-right scan, whereby the sum of the intensity values is listed and calculate the mode. The mode value serves as an indicator of the number of waveforms and the position of the listed value is indicative of the region in which the waveform is situated. These indices enable the assignment of height coordinates to waveforms by specifying the intersection points, allowing the waveform to be extracted in the form of slices. Furthermore, the utilization of identical coordinates enables the simultaneous extraction of the identical gridline and text areas.

Signal range estimation: ECG grid comprises two distinct of lines: vertical lines representing time and horizontal lines representing amplitude. Accordingly, the number of horizontal lines can be used to estimate the amplitude range of the signal. Algorithm [12] is employed to extract horizontal lines from the gridline channel and skeletonize in order to convert their thickness to 1. To reconnection of disrupted lines, dilation algorithm [13] we used. The number of horizontal lines, n, can be determined by mode value after scanning left-to-right through the kernel of size (*height*, *1*) in a manner analogous to that employed in detecting the waveform region. In an ECG image, the typical value of a small grid box is 0.1 millivolt (mV). Therefore, n is multiplied by 0.1 to determine the maximum signal range r of the waveform.

Converting to signal: In the waveform channel of the slice, the width is indicative of time, the height is representative of the signal range. To transform the waveform channel of slices into a one-dimensional signal, a scan is conducted on each slice from left to right using a kernel of size (*height, 1*) During this process, the height indices where non-zero values exist are listed, representing signal values. In the absence of a non-zero value at any point, a NaN (Not a Number) value is used. These NaN values are then interpolated and reconstructed a digitized signal s. The signal s is normalized to a range $(0, r)$ using min-max scaling. Subsequently, the signal range is approximated as an amplitude by subtracting the mean value of the normalized signal s.

3. Results

As demonstrated in Table 1, the final challenge scores and ranks achieved by our team for both classification and segmentation tasks are presented.

Task	Score	Rank
Digitization	SNR: 5.493	2/16
Classification	F-measure: 0.730	3/16

Table 1. The signal-to-noise ratio (SNR) and F-measure of our team's model on the hidden data [14] for the digitization and classification tasks.

4. Discussion and Conclusions

In this study, we proposed a method for extracting key components to digitize and classify 12-lead ECG images. The effectiveness of our approach is demonstrated by its performance across both digitization and classification tasks, which has practical implications. However, further refinement is necessary in several areas. The classification model employs an excessively large input data set, necessitating supplementary sizing experiments to rectify the inefficiency. Furthermore, the digitization algorithm operates as a rule-based algorithm, rendering performance susceptible to reductions when unexpected exceptions occur.

In conclusion, our proposed method demonstrates significant promise; however, there remains potential for further improvement, particularly in reducing computational costs and enhancing robustness to signal variability. Future research should focus on optimizing the classification input size and transitioning to a more adaptive deep learning model for signal conversion in the digitization algorithm.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which is funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR20C0026).

References

[1] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[2] M. A. Reyna, Deepanshi, J. Weigle, Z. Koscova, A. Elola, S. Seyedi, K. Campbell, G. D. Clifford, R. Sameni, “Digitization and classification of ECG images: The George B. Moody PhysioNet challenge 2024,” *Computing in Cardiology*, no. 51, pp. 1-4, 2024.

[3] P. Wagner, N. Strodthoff, R. Bousseljot, W. Samek, T. Schaeffter, “PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3),” *PhysioNet*, 2022.

[4] P. Wagner, N. Strodthoff, R. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, T. Schaeffter, “PTB-XL: A large publicly available ECG dataset,” *Scientific Data*, no. 154, 2020.

[5] E. A. Perez Alday, A. Gu, A. J. Shah, C. Robichaux, et al. “Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology challenge 2020,” *Physiological Measurement*, Nov. 2020.

[6] M. A. Reyna, N. Sadr, E. A. Perez Alday, A. Gu, et al. “Will two do? Varying dimensions in electrocardiography: The PhysioNet/Computing in Cardiology challenge 2021,” *Computing in Cardiology*, no. 48, pp. 1-4, 2021.

[7] B. Gow, T. Pollard, L. A. Nathanson, A. Johnson, B. Moody, et al. “MIMIC-IV-ECG: Diagnostic electrocardiogram matched subset (version 1.0),” *PhysioNet*, 2023.

[8] K. K. Shivashankara, Deepanshi, A. M. Shervedani, M. A. Reyna, G. D. Clifford, R. Sameni, “ECG-Image-Kit: A synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization,” *Physiological Measurement*, vol. 45, pp. 055019, 2024.

[9] Deepanshi, K. K. Shivashankara, G. D. Clifford, M. A. Reyna, R. Sameni, “ECG-Image-Kit: A toolkit for synthesis, analysis, and digitization of electrocardiogram images,” Jan. 2024.

[10] M. Tan, Q. V. Le, “EfficientNet: Rethinking model scaling or convolutional neural networks,” *ICML*, pp. 6105–6114, 2019.

[11] T. Y. Zhang, C. Y. Suen, “A fast parallel algorithm for thinning digital patterns,” *Communications of the ACM*, vol. 27, no. 3, pp.236–239, 1984.

[12] T. Tsesmelis, “Extract horizontal and vertical lines by using morphological operations,” *Open Source Computer Vision*, `https://docs.opencv.org/4.x/dd/dd7/tutorial_morph_lines_detection.html`.

[13] H. Heijmans and C. Ronse, “The algebraic basis of mathematical morphology: I. dilations and erosions,” *Computer Vision, Graphics, and Image Processing*, no.50, pp. 245–295, 1990.

[14] M. A. Reyna, Deepanshi, J. Weigle, Z. Koscova, K. Campbell, K. K. Shivashankara, 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,” 2024. URL `https://arxiv.org/abs/2409.16612`.

Address for correspondence:

Sung-Hoon Kim
88, Olympic-ro 43-gil, Songpa-gu, Seoul, Republic of Korea
shkimans@gmail.com