WAVIE: A Modular and Open-Source Python Implementation for Fully Automated Digitisation of Paper Electrocardiograms

Mathilde A Verlyck[1], Joshua R Dillon[1], Stephen A Creamer[1], Debbie Zhao[1]

[1] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Abstract

The electrocardiogram (ECG) is a ubiquitous tool for the assessment of heart disease. While considerable effort has been directed at ECG digitization to facilitate artificial intelligence applications, limitations persist in the generalizability of existing methods. As part of the 'Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024', we present WAVIE, a fully-automated, modular, and open-source framework for ECG digitization to handle the heterogeneity of real-world data. Using the PTB-XL dataset, synthetic paper ECGs were generated with known variations and artifacts. Our team, wavie_ABI, developed a three-stage framework consisting of deep-learning models for orientation correction, object detection, and waveform extraction. Inference on the hidden test set for the digitisation task produced a mean signal-to-noise ratio (SNR) of 5.469 (ranked 3rd of 16 teams). WAVIE provides a comprehensive and generalizable baseline that can be reconfigured and fine-tuned for specific ECG digitisation tasks, ensuring adaptability for future research applications.

1. Introduction

The electrocardiogram (ECG) is a frontline test for rapid and non-invasive assessment of the electrical activity of the heart. In practice, the majority of ECGs are stored as paper records or captured as images, which are subsequently manually interpreted by a physician. A large amount of data has been acquired to date, which may be leveraged for the development of computer-aided approaches for ECG interpretation [1]. Such advancements have the potential to improve ECG-based cardiovascular disease prediction and facilitate widespread screening.

In a real-world scenario, the storage of ECGs in paper or image formats can result in various artifacts such as creasing, stains, or the addition of pen marks, which may distort or obscure the underlying signal of interest. Furthermore, the layout of ECG data is not necessarily uniform across all systems, producing inconsistencies in various elements such as grid size, colouring, padding, and the presence of additional headers or patient information. Consequently, considerable effort has been directed at extracting digital waveforms from paper-based records in order to enable automated analyses and the application of modern artificial intelligence techniques for disease prediction.

To overcome these challenges and limitations with existing tools, the George B. Moody PhysioNet Challenge 2024 provides a platform to advance the field of ECG-based diagnosis by inviting teams to digitize and classify ECGs captured from images or paper printouts [2, 3]. In contribution to this challenge, we present WAVIE, a fully automated, modular, and Python-based framework for ECG digitization that can handle the heterogeneity of real-world ECG formats.

2. Methods

This section describes the processes relevant to synthetic ECG data generation and the application of several established open-source deep learning models to carry out various tasks related to digitization. Figure 1 provides an overview of the modules within the WAVIE framework.

2.1. Data Preparation

The PTB-XL [4, 5] dataset consisting of 21799 records of 12-lead ECG waveform was used to train our model. A held-out subset of the dataset was randomly created using a 90/10 split (19628/2170) (one record was excluded due to a metadata failure).

Synthetic paper ECGs with known waveforms and lead-specific bounding boxes were generated through the use of the synthetic image generation toolbox, ECG-Image-Kit [6, 7]. To validate the methods on a broad range of variations and artifacts, five different generation configurations (shown in Table 1) were used for the training data and ten for the held-out data. Six of the ten held-out configurations were unseen, allowing for the evaluation of model performance on novel data (containing new records, noise, and artifacts) to assess generalizability.

Two minor modifications were made to the ECG-Image-

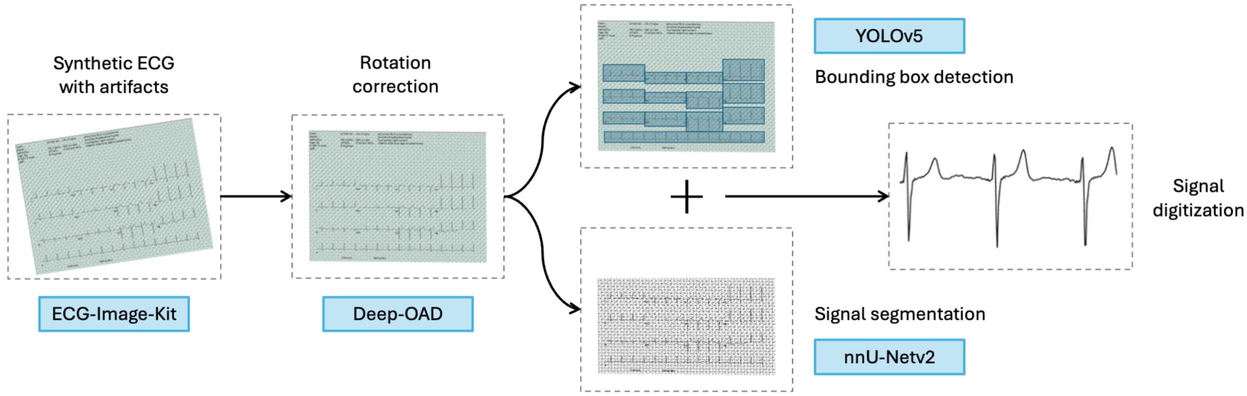

Figure 1. Overview of WAVIE framework consisting of open source modules for ECG digitization from natural and synthetic paper ECGs.

Kit source code. First, a condition was added to the rhythm lead selection process. If the selected rhythm lead was less than 10 seconds, the default rhythm lead ('II') was used. Second, it was ensured that two-column images were only generated in the case where each signal lead was at least 5 seconds long. These conditions prevented the creation of unrealistic partial signals on synthetic paper ECGs. It is also of note that no rotations were applied using ECG-image-Kit. Instead, a subset of 113 images from the held-out set (~5%) was rotated externally between -30 and 30 degrees to simulate a realistic range of rotations and with random border colours.

A preprocessing step handled image conversion to .JPEG and created a training-validation split for the rotation and bounding box detection models. The training set described previously was further divided into training and validation sets with a ratio of 80/20, resulting in 15702 training images and 3926 validation images. The lead bounding boxes were padded vertically by ±6 pixels to capture the entirety of the signal.

Ground truth labels for the segmentation model were prepared by creating a binary mask of the image. Each lead signal was masked using the plotted pixel field generated by ECG-Image-Kit and overlaid onto the full image. The mask was then dilated with a 2x2 square kernel to preserve peak morphology.

2.2. Digitization Modules

2.2.1. Rotation correction

Deep Image Orientation Angle Detection [8] is a deep neural network designed to estimate the rotation angle of natural images. During training, the images are rotated on the fly, and the network predicts the estimated angle of rotation. We modified the rotation distribution from the default uniform sampling between 0 and 360 degrees to normal distribution sampling, returning a higher number of samples at low angles ($\mu = 0$, $\sigma = 25.2$).

Due to the hypothesis that real-world images were more likely to be scanned with smaller rotation angles, with the observation that the model otherwise tended to overfit on larger rotation angles. The inverse of the predicted angle was subsequently applied to the image to align it, sometimes resulting in signal cropping. This step ensured that the next modules were trained on images with a standardized orientation.

2.2.2. Signal segmentation

nnU-Net is an open-source self-configuring U-Net optimised for medical image segmentation [9]. A major advantage of nnU-Net is that network topology and hyperparameter selection are automatically chosen based on the characteristics of the training data. We trained an out-of-the-box 2D nnU-Net for 250 epochs, and an initial learning rate of 0.01 decreased according to the polyLR schedule [10]. On-the-fly augmentations, including rotation, downsampling, and mirroring, were carried out to improve model robustness. The trained model took in the output of the rotation network and returned a mask containing all lead signals in the ECG image.

2.2.3. Detection of lead bounding boxes

YOLO (You Only Look Once) [11] is a well-established, open-source model for end-to-end object detection and classification. YOLOv5 [12], proposes a good trade-off between accuracy and complexity and is implemented in Pytorch, a well-documented open-source framework for

Parameter	Train 1	Train 2	Train 3	Train 4	Train 5	Held-out 1	Held-out 2	Held-out 3	Held-out 4	Held-out 5	Held-out 6	Held-out 7	Held-out 8	Held-out 9	Held-out 10
Random padding	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	-	TRUE	-	TRUE
Pad inches	2	2	2	2	2	2	2	2	2	2	2	-	2	-	2
Random print header	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	-	0.5
Random grid colour	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	-	TRUE
Random resolution	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	-	-	TRUE
Resolution	200	200	200	200	200	200	200	200	200	200	200	200	40	-	200
Random grid present	1	1	0	1	1	1	1	0	1	1	0	1	1	-	1
Calibration pulse	0.5	1	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	-	1
Number of columns	4	4	4	4	4	2	4	4	4	4	4	4	4	-	4
Add QR code	-	-	-	-	TRUE	-	-	-	-	TRUE	-	-	-	-	-
Full mode	-	-	-	V1	-	-	-	-	V1	-	-	-	-	-	I
Wrinkles	-	TRUE	TRUE	TRUE	TRUE	-	TRUE	TRUE	TRUE	TRUE	TRUE	-	-	-	-
Crease angle	-	45	75	135	90	-	45	75	135	90	135	-	-	-	-
No of vertical creases	-	15	5	20	10	-	15	5	20	10	20	-	-	-	-
No of horizontal creases	-	15	5	20	10	-	15	5	20	10	20	-	-	-	-
Augment	-	TRUE	TRUE	TRUE	TRUE	-	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	-	-
Noise level	-	37	25	12	50	-	37	25	12	50	50	15	5	-	-
Cropping	-	-	-	-	-	-	-	-	-	-	-	0.08	-	-	-
Handwritten text	-	-	-	-	TRUE	-	-	-	-	TRUE	-	-	-	-	TRUE
No of words to add	-	-	-	-	10	-	-	-	-	10	-	-	-	-	10
X offset	-	-	-	-	400	-	-	-	-	400	-	-	-	-	400
Y offset	-	-	-	-	400	-	-	-	-	400	-	-	-	-	400

Table 1. Parameters used for ECG-Image-Kit for generation of synthetic ECG images for training and held-out sets.

machine learning applications. YOLOv5 applies augmentation techniques such as random flipping, affine transformation, and mosaic augmentation to improve generalizability. We trained the model for 300 epochs using the pre-trained weights of YOLOv5s (YOLOv5 implemented with small architecture) and the optimized YOLOv5s hyperparameters of the pre-trained model [12]. The small architecture was selected for its low computational resource demand and fast inference times. The ground truth consisted of lead bounding boxes surrounding the signals and bounding boxes for the lead names to detect the lead name class. A post-processing step was implemented to ensure that only the 13 lead bounding boxes (12-lead ECG and one rhythm lead) were retained based on class confidence, overlap, and location, ensuring a 4×3 layout. The rhythm lead name was assigned using the closest lead name bounding box to the rhythm lead bounding box. The other lead names were assigned according to the ECG 4×3 template. Finally, the lead bounding boxes were converted to their initial format (x_min, y_min, x_max, y_max).

2.2.4. Signal digitization

The bounding boxes predicted by YOLOv5 were used to separate the nnU-Net predicted signal mask into separate leads for digitization. For each lead, the signal mask was extracted from its bounding box and contoured using OpenCV [13]. After removing extraneous contours, the median of the signal contour set at each column in the mask was taken as the signal position and converted into a voltage and time value (scaled using the standard ratio of 0.1mV/0.04s). The baseline position within each mask was approximated using a sliding window approach with the baseline set as the centre of the window with maximum standard deviation. The signal was then interpolated to the target length using a Piecewise Cubic Hermite Polynomial (PCHIP) interpolation scheme (to preserve peak morphology) and smoothed using a 1D cubic spline with a smoothing coefficient of 0.1.

3. Results

For the ECG digitisation task, our selected entry achieved an SNR of 5.469 on the hidden test set of the official phase of the challenge [14], placing our submission in third place as shown in Table 2. On our held-out set, we achieved an SNR of 1.676.

Task	Score	Ranking
Digitization	SNR: 5.469	3/16

Table 2. Signal-to-noise ratio (SNR) for our selected entry (team wavie_ABI) on the digitization task, including the ranking of our team on the hidden test set.

The challenge's evaluation code provided four additional metrics: the median SNR, the KS metric (a met-

	SNR	mSNR	KS	ASCI	WAD
Shift	1.676	7.857	0.936	-0.229	0.078
No shift	0.431	3.846	0.880	-0.846	0.123

Table 3. Additional metrics calculated on our held-out set with and without shift correction. SNR: signal-to-noise ratio, mSNR: median SNR, KS: Kolmogorov-Smirnov inspired metric, ASCI: adaptive signed correlation index, WAD: weighted absolute difference.

ric inspired by the Kolmogorov-Smirnov test statistic), the adaptive signed correlation index (ASCI) and the weighted absolute difference (WAD). These metrics were calculated on our held-out set with and without shift correction. The results are shown in Table 3. Furthermore, the digitization of one case took 30 seconds (wall clock).

4. Discussion and Conclusions

Our submission exhibited good results in terms of SNR, median SNR, WAD, and KS metric with shift correction, demonstrating the performance of WAVIE assessed using metrics related to both absolute errors and shape similarity of the reconstructed and reference waveform. Scores were decreased when shift correction was not applied due to the reduction in point-to-point correspondence, which is assumed to be strong in these metrics. However, strong results are still seen with respect to the KS metric, returning a value of 0.88 and the WAD, returning 0.123; combined, these show strong levels of shape similarity in the signals and errors of less than 13% in actual value.

A benefit of the modular nature of this workflow is demonstrated by the difference between shift-corrected and non-shift-corrected performance. This highlights the potential for a module that can accurately correct for distortions in vertical and horizontal shifting. Furthermore, compared to other solutions that are provided on an as-is basis, WAVIE can be readily reconfigured to meet project-specific requirements by incorporating modules to address dataset-related characteristics.

Future work will explore weaknesses and potential failure modes within the individual modules and address them through additional data augmentation, optimization of feature spaces, and consideration of additional modules when appropriate.

References

[1] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W, Schön TB, Ribeiro ALP. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature Communications April 2020;11(1). ISSN 2041-1723.

[2] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[3] Reyna MA, Deepanshi, Weigle J, Koscova Z, Elola A, Seyedi S, Campbell K, Clifford GD, Sameni R. Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024. Computing in Cardiology 2024;51:1–4.

[4] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data 2020; 7:154.

[5] Strodthoff N, Mehari T, Nagel C, Aston PJ, Sundar A, Graff C, Kanters JK, Haverkamp W, Dössel O, Loewe A, Bär M, Schaeffter T. PTB-XL+, a comprehensive electrocardiographic feature dataset. Scientific Data 2023;10:279.

[6] Shivashankara KK, Deepanshi, Shervedani AM, Reyna MA, Clifford GD, Sameni R. ECG-Image-Kit: a synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. Physiological Measurement 2024;45:055019.

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

[8] Maji S, Bose S. Deep image orientation angle detection. arXiv preprint arXiv200706709 2020;.

[9] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods December 2020;18(2):203–211. ISSN 1548-7105.

[10] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence April 2018;40(4):834–848. ISSN 2160-9292.

[11] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 2016; .

[12] Jocher G. YOLOv5 by Ultralytics, May 2020. URL `https://github.com/ultralytics/yolov5`.

[13] Itseez. Open source computer vision library. `https://github.com/itseez/opencv`, 2024.

[14] Reyna MA, Deepanshi, Weigle J, Koscova Z, Campbell K, Shivashankara KK, Saghafi S, Nikookar S, Motie-Shirazi M, Kiarashi Y, Seyedi S, Clifford GD, Sameni R. 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:

Mathilde A Verlyck
Auckland Bioengineering House, 70 Symonds Street, Grafton, Auckland 1010, New Zealand
mathilde.verlyck@auckland.ac.nz