YOUR-Lead: YOLO and U-Net for Reconstruction of ECG Lead Signals

Rafael Silva†, Yingyu Yang†, Maëlis Morier, Safaa Al-Ali, Maxime Sermesant

Centre Inria d'Université Côte d'Azur, Epione team, Sophia Antipolis, France

Abstract

Electrocardiograms (ECGs) are still vastly stored on paper, which restrains the development of data-driven algorithms for their analysis. We propose in this manuscript an efficient method to digitize ECGs from paper images, as part of the "Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024". A total of 21,799 12-lead recordings from the PTB-XL public data-set were used to generate realistic images. The architecture of our method consists of three main steps. First, we generate ECG paper images with different characteristics such as resolutions, layouts, grid patterns, wrinkles, temperatures, text and noise levels to reproduce real-life challenges. Second, we perform a fully automated ECG image processing pipeline. To do so, a U-Net model is employed for image binarization where only the relevant part of the image is extracted. In addition, a pre-trained YOLO object detection model is fine-tuned to detect the signal box leads. Combining the obtained cleaned ECG image and detected lead boxes, we finally extract 1D ECG signal for each lead present in the ECG paper. By comparing original and estimated ECG signals, our best model (Inria Epione team) achieved a signal-to-noise ratio (SNR) score of 0.558 on the hidden validation set (no ranking).

1. Introduction

The Electrocardiogram (ECG) corresponds to the recording of the cardiac electrical activity using body-surface electrodes. Multiple signals, also referred to as leads, are measured simultaneously as electrical potential differences between electrodes, which are placed in predefined locations of the body. The ECG remains a primary modality for monitoring cardiac events and diagnosing cardiovascular diseases, largely due to its widespread availability and ease of use. In healthcare facilities, the ECG is often printed on graph paper for visual analysis and storage, even though it can be easily exported and stored in digital format nowadays. Moreover, paper ECGs are prone to degradation and loss, and can represent a data security risk. Also, even with scanned versions of the ECG paper, raw signal data is not easily extracted.

†These authors contributed equally to this work

Certain cardiac abnormalities can be subtle and diagnosis heavily relies on the cardiologist level of expertise for accurate identification. In this context, data-driven analysis of ECG signals has the potential to yield more objective and robust insights into underlying cardiac conditions by extracting clinically relevant markers from the data [1]. Furthermore, automation through machine learning algorithms can streamline tasks that do not require expert interpretation. Since ECGs are intrinsically heterogeneous, automated data analysis can be performed consistently and be used to track subtle changes and trends over time, which can be challenging with manual analysis.

Therefore, there is an increased interest in converting ECG papers into a machine-readable format, especially when designing clinical retrospective studies. This conversion process known as digitization involves extracting ECG lead data points from scanned images and representing them as one-dimensional signals (i.e., as time series). However, this task can be difficult due to the diverse paper characteristics (such as layout, annotations, and grid) and image quality aspects (including noise, artifacts, and resolution). Therefore, effective image denoising becomes crucial to ensure accurate lead extraction. By leveraging the use of generated realistic images, it is possible to enhance the quality of denoising algorithms and facilitate the use of image processing tools meant for digitization.

As part of the George B. Moody PhysioNet Challenge 2024 [2, 3], we propose a methodology for ECG signal digitization that involves image denoising and lead box detection to identify and reconstruct the original ECG signals from scanned images.

2. Proposed approach

Our proposed method consists of three stages. First, we generate a realistic synthetic ECG paper data-set that includes ECG paper images and their binarized counterparts. Second, we design a fully automated pipeline to process the ECG paper. In this step, we train an object detection model to detect the position of separate lead signals and a denoising model to generate clean ECG images with only

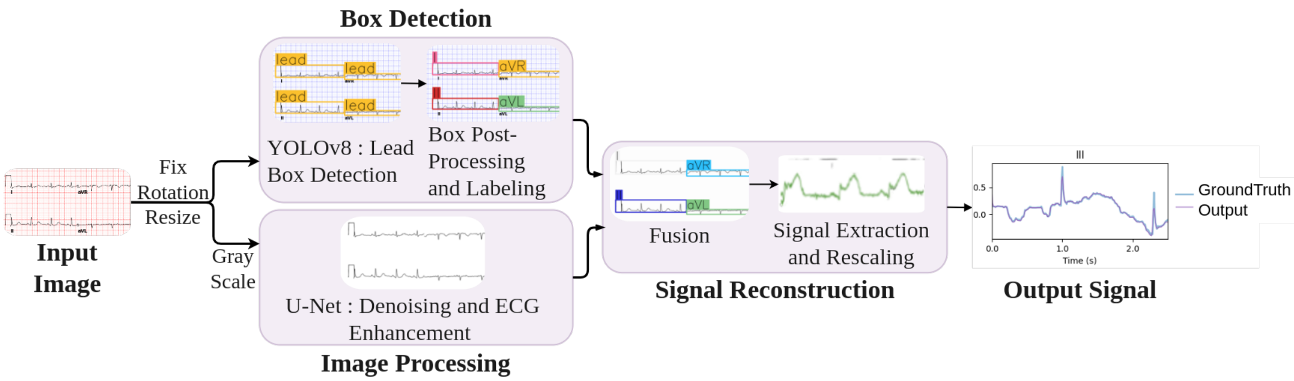

Figure 1. Workflow of YOUR-Lead for a given ECG paper image (we only show part of a 12-lead ECG image for better visualization). The input image patches pass through two image processing steps to obtain the corresponding binarized clean image and the lead boxes. The ECG signals are then reconstructed.

the signal traces. Finally, by combining the information of lead position and the cleaned ECG image, we extract 1D ECG signal for each lead present in the ECG paper. We depict the workflow of our proposed method in Fig. 1.

2.1. Data generation

A total of 21,799 ECG recordings each containing 12-leads information from the PTB-XL data-set [4] were used to generate ECG paper images that will be employed to train our model. In order to encompass a wide range of scenarios often encountered in clinical practice, we generate images with a variety of different characteristics, using the ECG-Image-Kit package [5,6]. These image properties include different resolutions, layouts, grid patterns, font styles, line widths, presence of text etc. In addition, we simulate realistic conditions of real scans such as wrinkles, temperature fluctuations, image blurring, color registration errors, aliasing, and sensor noise. With this data, the model used for ECG digitization can then be trained more efficiently, improving its ability to generalize and perform reliably under a wide range of conditions. In Fig. 2, we show examples of generated images with some artifacts: color temperature, paper wrinkles, Gaussian noise etc.

For each generated image, we extract the coordinates of the lead bounding boxes and also generate a gray-scale version containing only the signals, that is, with no grid, annotations, or artifacts. Along with the original generated image, these two elements will be used to perform automated image processing.

2.2. Image processing

Typically, an ECG image is acquired by scanning or photographing the graph paper printed by ECG machines. This image generally contains 12-leads overlaid on a standard grid background. To eliminate any rotation that might occur during scanning or photographing, we apply a rotation correction to ensure horizontal alignment of ECG images. We then train two models separately: one to generate a clean version of the ECG image containing only the signal traces, and the other one to detect the positions of the individual lead signals within the image.

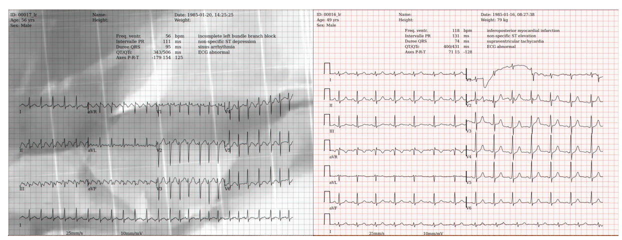

Figure 2. Examples of randomly generated ECG papers.

Rotation correction

A well-aligned ECG image exhibits recurrent patterns along both the x and y axes. In the frequency domain, this is characterized by dominant frequencies that are spread vertically and horizontally, intersecting at the origin. If an ECG image is rotated relative to its well-aligned version, the dominant frequency axes will also be rotated. We identify the rotation angle by analyzing these frequency components and then apply rotation correction to the original input image.

Image binarization

The `U-Net` convolutional neural network [7], widely used for image segmentation and document binarization, has been already employed for ECG extraction and detection [8]. For this challenge, this architecture is used to binarize the generated images, where only the relevant part of the image is extracted. To do so, the input gray-scaled images are first resized to 768×960 pixels, then split into patches of 192×192 each, which are then fed to the model. The U-Net architecture consists of 5 encoding blocks of 3×3 convolutional layers with [64, 128, 192, 512 and 1024] filters and 2×2 pooling layers, along with the corresponding decoding blocks and skip connections [7]. The training strategy uses the generated noisy image as input and the generated binary counterpart as target.

Lead box detection

We fine-tuned a pre-trained `YOLOv8` [9] model using the generated ECG paper data-set to perform the lead box detection. YOLO is a real-time object detection CNN that offers high detection accuracy and speed. YOLO will predict the bounding boxes of all present leads in the ECG paper, but we further need to filter the detected boxes and associate the appropriate lead names. Under the assumption of a 12-lead ECG paper, we identify the layout of the principal ECG lead signals (1×12, 2×6, 4×3) by filtering the median width of predicted lead boxes and regroup them into different columns using the KMeans clustering algorithm. Some additional post-processing techniques such as generating boxes that are missing based on the layout, re-aligning the vertical limits of the boxes, and correcting aberrant box position (Fig. 3) have been also carried out.

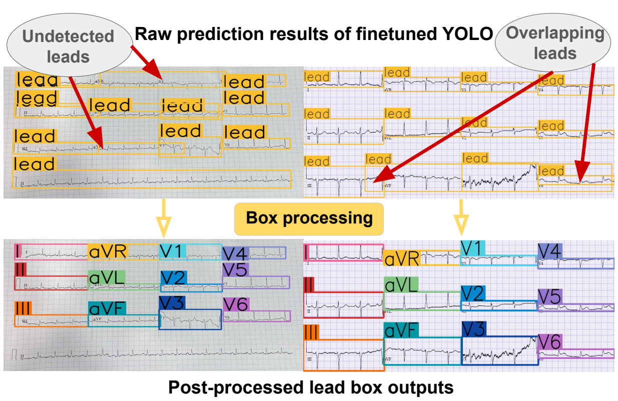

Figure 3. Examples of post-processing YOLO raw outputs to generate plausible 12-lead boxes.

2.3. Signal extraction and scaling

By combining the outputs of the image processing steps, we obtain a clean gray-scale image with the corresponding lead bounding boxes. From these boxes, the signals are extracted by reading the rows in each column that correspond to the signal. Since the bounding boxes can be contaminated with other ECG leads and artifacts, the connected components algorithm from the `OpenCV` library is used to isolate the pixels belonging to the lead signal. This is performed by first binarizing the box image and then selecting the connected component whose pixels are closest to the median of the box's darkest pixels, approximately corresponding to the isoelectric line. However, crossing ECG traces are still not eliminated with this approach. To minimize their effect, the row-wise median of the columns with the darkest pixels is picked. In cases of finding two distinct groups of pixels in one column, the rows closer to the previous signal value are chosen.

To scale the signals to their original units of acquisition, the paper's millimetric grid spacing needs to be estimated in pixels. We adapted the algorithm from [10] by first subtracting the binarized image to the original one to obtain a grid-enhanced image, and then finding the first peak of the auto-correlation signal, corresponding to the grid spacing. Given the standard vertical scale of 10 mm/mV and the horizontal scale of 25 mm/s used in medical-grade ECG papers, the extracted signals are appropriately scaled and interpolated.

3. Implementation

We fine-tuned a `YOLOv8n` model for 200 epochs using the default training settings. The target objects were the 12-lead signals, with the objective of detecting the positions of the 12-lead signal traces respectively in a given ECG image. We randomly split the generated data-set into 80% training, 10% validation, 10% testing.

To train the U-Net model, we used the AdamW optimizer [11] to update the learning rate, initially set to 1.17×10^{-3}. Since the target values are constrained between zero and one, we employed PyTorch's `BCEWithLogitsLoss` function. The images were processed in patches of size 192×192, in batches of 3 images. Training was conducted in Python 3.10, using PyTorch 2 framework, over 15 epochs. Bayesian hyperparameter optimization was used to find the optimal learning rate exploring a range from 1×10^{-5} to 1×10^{-1}, as well as the batch size, which was varied between 2 and 8.

4. Results and Discussion

To assess our model's effectiveness in the ECG digitizing task, the SNR was used. SNR is defined as the ratio between the power of the ground-truth signal and the power of the difference of the original and the reconstructed signals. We used 2000 held-out training samples to compute the training score (SNR -0.696), and present the score on the hidden validation set of the challenge in Table 1. The model scored 0.447 dB on the hidden validation data-set, slightly lower than our best entry with a minor modification (Table 1), indicating its capacity to correctly extract the lead signals. The detailed description of the data-set used for the challenge is described in [12].

Even though our model's SNR score is satisfactory, there are still a few areas where it may be improved to further improve its performance. For instance, there are cases where lead box detection fails, which may have an impact on the precision of signal extraction. Also, the binarization process is not always perfect, since there are situations where there are still artifacts present, and can distort the signal extraction process. Thus, the U-Net model might gain from being trained on a wider variety of noise and artifacts. Finally, our existing approach uses the median of the pixel values to estimate data points when ECG leads cross each other. Though generally effective, this method occasionally underestimates the QRS amplitudes, resulting in less accurate signal reconstruction.

Task	Score	Rank
Digitization	0.558	NaN

Table 1. SNR scores of our best entry (team Inria Epione) on the digitization task. The current method scored 0.447.

5. Conclusion

This paper presents an efficient three-step model to digitize ECG signals from paper images as part of "The George B. Moody PhysioNet Challenge 2024". Using ECG recordings from the PTB-XL data-set, we generated realistic ECG paper images. Our approach involves using a U-Net model for image binarization and a fine-tuned YOLO model for detecting the regions occupied by each lead. While our model achieves a satisfactory SNR score of 0.447 on the hidden validation set, there are still areas for improvement, namely on lead box detection and signal extraction in cases where ECG leads cross each other.

Acknowledgments and Funding

This work was supported by the SimCardioTest project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101016496 and by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. The authors thank the OPAL infrastructure from Université Côte d'Azur for providing resources and support.

References

[1] Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational Techniques for ECG Analysis and Interpretation in Light of Their Contribution to Medical Advances. Journal of The Royal Society Interface 2018; 15(138):20170821.

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

[3] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng C, Stanley H. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[4] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a Large Publicly Available Electrocardiography Dataset (Version 1.0.3), 2022. URL `https://doi.org/10.13026/kfzx-aw45`.

[5] Shivashankara KK, Shervedani AM, Clifford GD, Reyna MA, Sameni R, et al. ECG-Image-Kit: A Toolkit for Synthesis, Analysis, and Digitization of Electrocardiogram Images. Physiological Measurement 2024;45(5):055019.

[6] Deepanshi, Shivashankara K, Clifford G, Reyna M, 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`.

[7] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015; 234–241.

[8] Li Y, Qu Q, Wang M, Yu L, Wang J, Shen L, He K. Deep Learning for Digitizing Highly Noisy Paper-Based ECG Records. Computers in Biology and Medicine 2020; 127:104077. ISSN 0010-4825.

[9] 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. 2016; 779–788.

[10] Fortune JD, Coppa NE, Haq KT, Patel H, Tereshchenko LG. Digitizing ECG Image: A New Method and Open-Source Software Code. Computer Methods and Programs in Biomedicine 2022;221:106890. ISSN 0169-2607.

[11] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. In International Conference on Learning Representations. 2017; .

[12] 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:

Rafael Silva, Yingyu Yang
Centre INRIA d'Université Côte d'Azur, EPIONE team, 06902 Sophia Antipolis, France.
rafael.silva@inria.fr, yingyu.yang@inria.fr