A modular framework for the interpretation of paper ECGs

Sara Summerton[1], Nicola Dinsdale[2], Tuija Leinonen[3], George Searle[4], Matti Kaisti[3], David C Wong[5]

[1] University of Manchester, Manchester, UK
[2] University of Oxford, Oxford, UK [3] University of Turku, Turku, Finland [4] University College London, London, UK [5] University of Leeds, Leeds, UK

Abstract

Despite advances in digital storage of electrocardiograms (ECGs), paper print outs are still common place in clinical practice. The digitization and interpretation of paper ECGs is therefore of high utility. We describe the creation of a modular pipeline to achieve both of these tasks. The solution was created by the Easy Geese for the Digitization and Classification of ECG Images: George B. Moody PhysioNet Challenge 2024.

Methods: The pipeline accepts an image of a 12-lead ECG in any common format. It first extracts the area of interest using YOLO, and then segments pixels that constitute the ECG signals using a ResUnet. The resulting mask is rotated, and contiguous signal pixels are joined within the area of interest. In the last part of digitization, the signals are scaled, separated by lead, and checked for errors. Finally, the digitized 12-lead signals are input into an SE-resnet classifier to provide clinical interpretation.

Results: Our ResUnet had a Dice score of 0.997. On a subset of the training set, our digitization pipeline had an average signal-to-noise ratio (SNR) of 9.05; our ECG classifier had a macro f-measure of 0.70 ± 0.02*. This entry was unable to be scored on the validation set.*

1 Introduction

Despite the prevalence of large digital ECG databases for reseach, paper ECGs remain a stalwart of standard clinical care. While there have been multiple attempts to digitize images of paper ECGs, often these require user input to determine convert the raw signal into a clinically meaningful format. For instance, Fortune et al.'s approach requires a user to identify each ECG lead manually [1], while Santamonica's solution requires the location of reference pulses and the lead for any rhythm strips [2].

The 2024 George B. Moody PhysioNet Challenge was to digitize and classify ECGs directly from images, without human intervention. The task and dataset are described in detail in [3] and are provided via Physionet [4].

2 Methods

Training images for this challenge were generated from signals from the PTB-XL database using the ecg-image-kit generator [5,6]. The generator provided images in a format similar to a paper ECG. The generator was also modified to provide an ECG image mask – ECG signals with no grid, text, or other augmentation. Clinical labels were available directly from PTB-XL. Labels were pre-processed into a reduced set of 11 classes for the challenge.

We created additional training examples in two ways. First, we printed these generated images on paper and scanned them to recreate a digital copy. We created multiple digital copies with physical augmentations including wrinkling and writing on the paper, and taking photos of the paper from different angles. Second, we sourced additional copyright-free scans of paper ECG images from the internet (`https://www.ecgguru.com/`). For a small subset of these images, we also created ECG masks manually by masking the ECG signal in image editing software.

We developed a modular pipeline that takes in a raw image and outputs a digitized 12-lead ECG signal and classification of the signal. An overview of the pipeline is shown in Figure 1 and is described in more detail below.

2.1 A. Determine number of ECG rows; B. Image Classification

We used a fine-tuned Yolov7 [7] to predict the bounding boxes of where individual ECG leads are present in an image. We use the bounding boxes to crop the image to the useful area and to determine the number of rows of ECG signal, which is used in *E. Pixel Joining*. We also classified the image as real or generated using a ResNet-18 model [8]. This step produced a binary flag that was used in the subsequent steps.

2.2 C. ECG segmentation

The task of binarising the ECG image was treated as a two class segmentation problem, and approached follow-

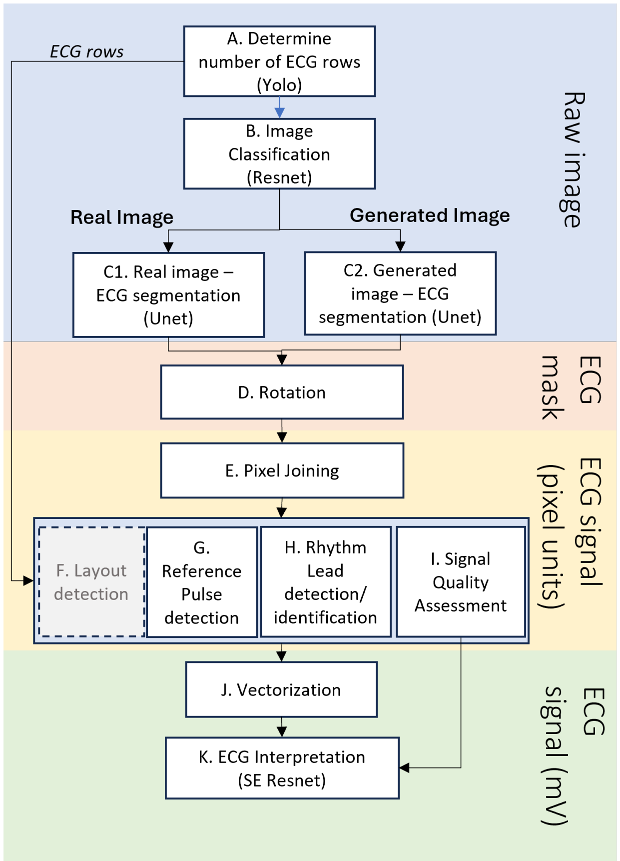

Figure 1. ECG digitization and interpretation pipeline.

ing the ResUNet method proposed by Li et al. [9], with the blocks at each stage formed of a 3 × 3 convolution, batch normalization and LeakyReLU non-linearity. The model was trained using a combination of dice and focal loss.

The ResUNet was trained on 4000 256×256 px patches. The model was also trained with augmentation designed to simulate the range of variation likely in the test set: colour jitters, blurring, random noise, rotations and scaling.

Given the very different characteristics of the real and simulated ECG images, it was not possible to produce a single model that performed well across both domains. Therefore, separate domain-specific models were trained for the real and simulated data respectively.

2.3 D. Rotation

We considered the simplified scenario in which raw ECG images could be rotated, but not subject to a more general perspective transform. Our solution assumed that the output mask of *C. ECG segmentation* was mostly accurate. The optimum rotation of the mask was that which minimized the number of columns containing ECG pixels.

2.4 E. Pixel Joining

The historical approach to ECG digitization, as implemented by [2] among others, assumes that an image contains a contiguous selection of pixels that correspond to an ECG signal, and that at least one ECG pixel is present per column of the image. Dynamic programming is then employed to find a minimum cost path across the image. This approach can fail when the image is noisy and when ECG signals overlap when adjacent leads have large amplitudes.

We developed an alternative simple forward-search, depicted in Figure 2. It takes a binary Unet output mask with labels background, ECG as input. We assume that we know the coordinates of an initial ECG pixel in the left-most column (a), and that we have a rough estimate of the baseline (red line). From the initial pixel, all contiguous ECG pixels in the column selected (b). We then consider in the next column and select all ECG pixel with minimum manhattan distance to the existing set of ECG pixels (c). If more there is more than one candidate pixel, we select the pixel closest to the baseline. If there are no candidate pixels in the next column, then skip a column (d). This process is repeated until the end of the image. Finally, the pixel with maximum distance to the baseline is retained to produce the ECG signal in pixel units (e).

2.5 F-I. Lead reconstruction

F. Rhythm Lead detection and identification The rhythm lead is a single ECG lead that is recorded for a full ten seconds, and typically spans a whole ECG recording. Given the challenge assumption of a standard 3x4 grid layout of short ECG segments, any additional rows were assumed to be rhythm leads. We assumed that lead name of each rhythm strip was fully deterministic, depending on the total number of rhythm leads: lead II if one rhythm strip, leads II, V5 if two, and leads V1, II, V5 if three.

F. Layout detection Real ECG images can vary in both lead order and lead layout on an image. We used correlations between leads to detect the alternative Cabrera format for lead order. Given the limitations of the generator, this was only applied to real images.

G. Reference Pulse Detection The reference pulse is a square wave denoting 1 mV in amplitude and 0.2 seconds in duration. While it is commonly to the left of ECG, it can also appear to the right, or may not be present. Detection 1. informs ECG signal baseline and scaling parameters, and 2. ensures it is not mistaken for part of the ECG.

We detect the location of a reference pulse by searching for square waves of equal width in the same position in at least three lines of the digitized signal.

I. Signal Quality Assessment Challenges in reconstruction occurred when *E. Pixel Joining* failed to correctly identify the start or end of a signal, or returned the same

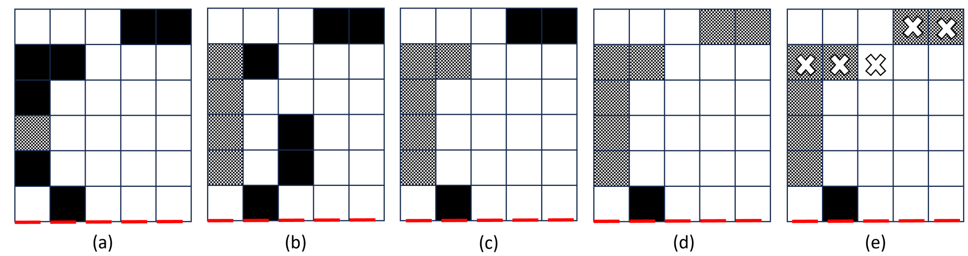

Figure 2. Pixel joining algorithm to extract ECG signal from u-net output mask. (a) Initially, any ECG pixel is selected in the left-most column. (b) All adjoining ECG pixels in the column are added. (c) ECG pixels in the next column are considered; pixel with minimum distance to previous column is added. (d) Where no candidate pixels exist, column is skipped. (e) One pixel per column is selected to generate the ECG signal.

set of overlapping pixels for a single signal. To mitigate the impact of errant reconstructions on overall SNR, we returned NaNs for a single lead that both 1. overlapped with another and 2. departed from its baseline for a prolonged period. Furthermore, if rows of the reconstructed signal had different temporal (length) scaling factors, we assumed pixel joining had not been completed successfully, and the entire reconstruction output was set to NaN.

2.6 J. Vectorization

The final digitization stage was to convert the ECG in pixel units into real units of mV and seconds. To do this, we first remove the reference pulse. We then scale the signal, using the provided duration of the ECG signal, and the fact that 0.2 s of length corresponds to 5 mV in amplitude (height). Finally, in instances where a reference pulse is available, we set the bottom of the reference pulse to be the baseline of 0 mV. Otherwise, we use the median value (in pixels) of a row of ECG to be 0 mV.

2.7 K. ECG Interpretation

We interpreted the resulting 12-lead signal using a modified version of Zhao et al's SE Resnet [10,11]. This model output a set of probabilities corresponding to 11 labels.

We retrained the model using data with missingness corresponding to a standard 12-lead recording (i.e. with 2.5 s for most leads, instead of the original 10 s). We trained the model three times under different starting weights, and ensembled the models using the mean probabilities. To account for partially successful digitization, we trained an additional model with missing leads. A flag from *I.* was used to switch between the two models at inference time.

To convert output probabilities to labels, we selected the label with the highest probability to be the primary class. Secondary classes were then selected if their probability exceeded a heuristic threshold of 0.3, determined by maximizing F-measure on a held-out subset of the training set.

Task	Training	Validation	Test	Rank
Digit.	9.05	N/A	N/A	N/A
Class.	0.70 ± 0.02	N/A	N/A	N/A

Table 1. Signal-to-noise ratio (SNR) for the digitization ('Digit.') task and macro F-measure for the classification ('Class.') task. Due to time constraints, SNR was averaged over 500 images with no noise or augmentations and 1000 images with noise, rotation, creases, and other artifacts, generated from a held-out subset of the training set. The ensemble classifier was evaluated using 5-fold cross-validation on data directly from PTB-XL.

3 Results

The U-net performed well on generated signals, with a Dice score of 0.997. An example input image with shadow, corresponding output mask with pixel joining, and the reconstructed signal are shown in Figure 3. From visual inspection, we concluded that the U-net performed worse on real images, likely due to the very limited training data.

Our scores are given in Table 1 We were unable to be scored on the official validation set, and note our local evaluation is not comparable to official challenge scores.

4 Discussion

We developed an analysis pipeline that automatically interprets 12-lead ECG from an image. Our approach assumed that digitization was necessary to retrieve then classify the informative signal.

One significant weakness of our approach was that errors in individual modules could compound. We were particularly reliant on the performance of the *ECG segmentation* step; minor errors here led to large changes in SNR. Another shortcoming of our approach was that it did not consider perspective transforms of an image, nor local non-linear deformation from paper creases. This was also a weakness of the generator used to construct training data. In future work, we intend to improve the pipeline to handle

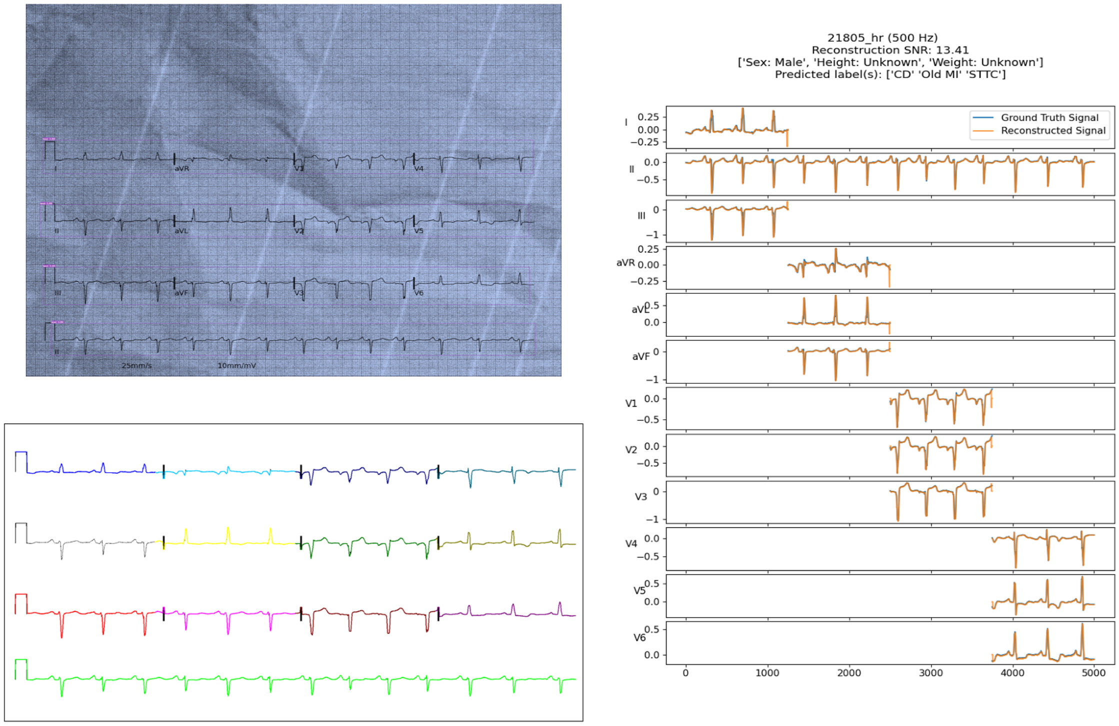

Figure 3. Example output of the digitization process (steps A-J). (a) shows the input image including noise and shadow, with detected areas of interest. (b) shows the unet output ECG mask (step C), and the output of *D. Pixel Joining* in colour. (c) shows the output signal in comparison to the PTB-XL original signal. Unplotted portions of the signal are not shown.

perspective transforms. We will also investigate imputation of unreconstructed data using cross-lead information.

References

[1] 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.

[2] Santamónica AF, Carratalá-Sáez R, Larriba Y, Pérez-Castellanos A, Rueda C. Ecgminer: A flexible software for accurately digitizing ecg. Computer methods and programs in biomedicine 2024;246:108053.

[3] 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. In 2024 Computing in Cardiology, volume 51. 2024; 1–4.

[4] 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.

[5] 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(1).

[6] Shivashankara KK, Clifford GD, Reyna MA, Sameni R. Ecg-image-kit: a synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. `https://github.com/alphanumericslab/ecg-image-kit`, 2024.

[7] Wang CY, Bochkovskiy A, Liao HYM. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023; 7464–7475.

[8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. CoRR 2015;abs/1512.03385.

[9] 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.

[10] Zhao Z, Fang H, Relton SD, Yan R, Liu Y, Li Z, Qin J, Wong DC. Adaptive lead weighted resnet trained with different duration signals for classifying 12-lead ecgs. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.

[11] Zhao Z, Murphy D, Gifford H, Williams S, Darlington A, Relton SD, Fang H, Wong DC. Analysis of an adaptive lead weighted resnet for multiclass classification of 12-lead ecgs. Physiological Measurement 2022;43(3):034001.

Address for correspondence:

Sara Summerton

Kilburn Building, Oxford Road, M13 9PL, Manchester, UK

sara.summerton@manchester.ac.uk