Dual Deep Learning System to Digitize and Classify 12-lead ECGs from Scanned Images

Chun-Ti Chou*, Sergio González*

AI Center, Inventec Corporation, Taipei, Taiwan

Abstract

As part of the PhysioNet/Computing in Cardiology Challenge 2024, our team, Inventec AIC, developed a dual deep-learning system to digitize and classify 12-lead electrocardiograms (ECG) from scanned images. Our approach comprises a computer vision (CV) algorithm and two deep-learning models based on Convolutional Neural Networks (CNN). Our preprocessing algorithm uses contour detection to capture the ECG grid, cropping out non-relevant information and fixing rotations. Our digitization approach leverages a fine-tuned object detection algorithm – YOLOv7 to detect and crop the different ECG sequences. Then, our digitization model based on CNNs with self-attention outputs the digital ECG signals. Besides, we developed an EfficientNet-B0 model with sample weighting to classify ECG images into 11 labels. We trained our models with 336K synthetic images with different formats and distortions from three datasets. In our internal evaluation, our approaches achieved an SNR of 1.479 and a macro F-measure of 0.728. For the digitization task, our model received an SNR of 0.397 (ranked 21th out of 81 submissions) on the hidden validation set. For the classification task, our model received a macro F-measure of 0.742 (ranked 5th of 90 submissions).

1. Introduction

Digital electrocardiogram (ECG) has been very beneficial for data integration, interpretation, and detection of cardiac diseases. However, ECG papers have remained common globally. Digitization and analysis of ECG papers can ensure comprehensive cardiac care across all demographics and regions. The 2024 George B. Moody PhysioNet Challenge invited teams to develop automated, open-source solutions to such a problem [1, 2]. We have designed a dual deep-learning approach combining a computer vision (CV) preprocessing algorithm and two Convolutional Neural Networks (CNN). Our digitization approach includes an fine-tuned YOLOv7 model [3] and a self-designed CNN model with a self-attention layer. Our classifier is an EfficientNet-B0 [4] with sample weighting to classify the ECG images into 11 cardiac abnormalities. We trained our models using synthetic ECG images with diverse formats and distortions generated by our modified version of ECG-Image-Kit [5, 6] from three different datasets [7–11].

2. Methods

2.1. Data generation and preprocessing

Our system uses models pre-trained with a large dataset of diverse synthetic ECG images. Thus, data generation and preprocessing are crucial for model generalization. Apart from PTB-XL [7,8], we have considered CODE15% [11] and the 2021 PhysioNet Challenge datasets [9, 10]. Each dataset has its characteristics and labeling criteria, needing data relabelling and signal filtering or removal. For CODE15%, we relabeled its classes by assigning AFIB, sinus bradycardia, sinus tachycardia, and normal ECG to their corresponding labels and AV block, RBBB, and LBBB to CD. Due to frequent noisy signals and baseline wander, we filtered the ECGs with a Butterworth bandpass filter of 5th order between 0.05 - 150 Hz. Besides, we dropped those records shorter than 10s and those without any positive label because the latter caused class-imbalanced distributions with redundant information. CODE15% dataset lacks some labels (Acute MI, HYP, Old MI, PAC, PVC, and STTC), which should be considered to avoid divergence during training. The 2021 PhysioNet Challenge datasets have more than 130 different cardiac abnormalities. We relabeled them into the challenge labels by checking their descriptions with their SNOMEDCT codes. Our mapping table is available online[1]. ECGs with only a sinus rhythm label were considered Normal ECGs. Besides, we filtered St Peterburg records similarly to CODE15%.

From approximately 223K records, we generated more

* Both authors contributed equally to this paper.

[1]Mapping for 2021 PhysioNet Challenge datasets: `https://drive.google.com/file/d/13T9kUe8Rh_MNa98T0jK_ZJ29ELyMrmii`.

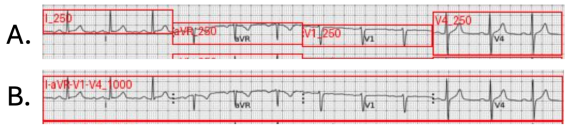

Figure 1. Original (A) and Merged (B) bounding boxes.

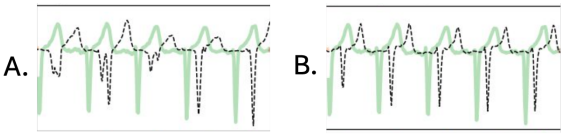

Figure 2. Models trained with MSE (A) and soft-DTW (B). Green lines represent the ground-truth ECG signal and dashed lines are the prediction.

than 336K synthetic ECG images with our extension of ECG-Image-Kit [5, 6]. In addition to the existing distortions, we varied the lead placement and grid style to resemble actual ECG papers and avoid positional biases. We included different 3x4 ECG configurations with none, 1 (II), 2 (V1, II), or 3 rhythm leads (V1, II, V5). We also changed from a fixed lead placement to distributing the space equidistantly or based on the signal amplitude. We randomized the location of the lead labels somewhere between the beginning and middle of the signal and above or below its baseline. Finally, we introduced different line styles for the grid and lead dividers.

Before training and inference, we designed a CV algorithm to detect the ECG grid, cropping out non-relevant information and fixing image rotations. First, the algorithm applies a gray-scale transformation, a 9-kernel Gaussian blurring, and an adaptive thresholding to the ECG image. Our algorithm then extracts the maximum area contour, fills it with a convex hull, and creates a binary image with the contour area. A second contour is detected on the binary image, fixing potential errors in the first detection. Afterwards, our algorithm extracts the corners of the minimum area rotated rectangle bounding the given mask. Finally, the image is cropped and rotated according to the corners.

To pre-train our models, we reserved 25% of PTB-XL for internal testing and splitting the other datasets and the remaining PTB-XL into 90%-10% for internal training and validation. The splitting was stratified to conserve the ratios of normal and abnormal ECGs for each dataset.

2.2. Digitization

The various formats of ECG papers make it challenging to digitize ECG signals using an end-to-end model. Thus, we proposed a two-stage method. In the detection stage, the YOLOv7 [3] detects the signals' location in the image. To fine-tune YOLO, we merged the bounding boxes in the same row, with the farthest edges as the new bounding box's boundaries as shown in Fig. 1. This merging process converts a 3x4 format into three 10-second signals so that the detection model can consistently focus on detecting ECG signals of uniform length, regardless of whether they originate from 3x4 or rhythm signals.

In the digitization stage, the detected signal regions are cropped, scaled to a predefined size, and passed through our digitization model. This model comprises 2D CNN blocks, self-attention layers, and 1D transpose CNN blocks. The 2D CNN blocks aim to encode the input image, and a global average pooling is applied along the height, keeping the time information. The self-attention and transpose CNN layers aim to aggregate information from each segment and generate the digital signals, respectively. Table 1 shows the detailed settings.

Mean Square Error (MSE) was initially used as the loss function. However, the model tended to produce extra pulses and noisy artifacts, as shown in Fig. 2. Consequently, we introduced Soft Dynamic Time Warping (soft-DTW) [12] as a secondary loss function. Unlike MSE, DTW calculates the error based on "paired" points, which can tolerate time misalignment and focus on morphology similarity. The final loss combined MSE and DTW losses with a hyperparameter α, set to 0.5 in our final model.

To enhance the robustness against noise in real images, augmentation methods, such as rotation and perspective transformation, were applied to the YOLO and digitization models.

2.3. Classification

We developed a multi-label end-to-end CNN model to classify mid-resolution ECG images into the 11 labels of the challenge. As backbone architecture, we chose an EfficientNet-B0 [4] after trying other image classification models and self-designed architectures. We set the default parameters in Table 1. However, we changed the original squared image size from 224 to 600 pixels because we believed the former resizing would remove relevant ECG features. We did not scale up the model to larger configurations such as B7 due to the GPU memory limitation.

During our training, some ECG records had multiple distorted images from the same 10s records as in PTB-XL or because they are longer than 10s as in CPSC and St Peterburg datasets. In such cases, we fed the model with a different image randomly selected from their pool per epoch. We also applied our CV algorithm described above to account for its errors and some data augmentations with the configurations shown in Table 1. We used binary cross-entropy loss with sample weighting to handle unlabelled samples and class imbalance. When training with CODE15% samples, we set weights to 0 for their

Digitization	
2D Residual Blocks	
Image size (H, W)	(150, 500)
Kernel sizes (H, W)	(3, 5)
Num. of blocks	4
Output channels	[32, 32, 64, 64]
Output length	[250, 250, 125, 125]
Transformer Layer	
Num. of attention head	4
Num. of Transformer Encoder(s)	2
1D Transposed Residual Blocks	
Kernel sizes	3 for CNN, 4 for transposed CNN
Num. of blocks	4
Output channels	[64, 32, 16, 8]
Output length	[125, 250, 500, 1000]
Classification	
EfficientNet-B0	
Image size	600
Width & Depth coefficients	1
Depth divisor	8
Kernel sizes	[3, 3, 5, 3, 5, 5, 3]
Input channels	[32, 16, 24, 40, 80, 112, 192]
Output channels	[16, 24, 40, 80, 112, 192, 320]
Strides	[1, 2, 2, 2, 1, 2, 1]
Num. of block repeats	[1, 2, 2, 3, 3, 4, 1]
Expand ratios	[1, 6, 6, 6, 6, 6, 6]
Squeeze expansion ratio	0.25
Activation function	'swish'
Hidden dimension	1280
Batch norm. eps. & momentum	0.001 & 0.99
Dropout rate	0.5 for CNN, 0.2 for FCN
Data augmentation	
Random affine	degrees: 15, scale: [0.9,1.1], translate: [0.05,0.15]
Random perspective	scale: 0.25, p: 0.3
Color Jitter	brightness: 0.1, contrast: 0.1 saturation: 0.1, hue: 0.05

Table 1. Hyperparameters for each model and task.

unlabeled classes, excluding them from loss propagation. For class imbalance, we assign weights w_j to the positive samples of a given class j equal to its number of negative samples divided by the maximum number of positive samples across all classes:

$$w_j = \frac{\sum_{i=0}^{N} 1 - y_{ij}}{\max_{j'} \sum_{i=0}^{N} y_{ij'}} \quad (1)$$

where y_{ij} means whether the sample i belong to class j (1) or not (0). Our optimizer was AdamW with a 0.001 learning rate and early stopping monitored by F-measure.

Before inference, we first apply our CV algorithm to the ECG image. Next, our model assigns the predicted classes according to their tuned probability thresholds, which were chosen based on the best F-measure during training. For records with multiple images, the logits predicted for each image are aggregated by the maximum for PAC and PVC and the mean for the remaining classes. We decided on these aggregation functions by testing their performance with the long records of CPSC and St Peterburg.

Training	Validation	Test	Ranking
1.479	0.397	-	21/81

Table 2. SNR for our selected entry (team Inventec AIC) on the digitization task, including the ranking of our team on the hidden test. We show the score on our training hold-out fold, the best scoring on the hidden validation set, and one-time scoring on the hidden test.

Unmodified	with Affine	Various formats
2.288	1.485	0.664

Table 3. SNR under different image configurations. "with Affine" includes rotation and perspective changes. "Various formats" contains arbitrary numbers of rhythm signals.

3. Results

Table 2 presents the digitization results on the training, validation, and test datasets. The SNR score on the validation set is 0.397, significantly lower than the score on the training set. Table 3 displays the evaluation results on different image configurations, revealing that affine transformations severely degrade the performance of the digitization model. Additionally, the method struggles with various paper formats, possibly because the images for training were biased in some specific formats.

Additionally, a noticeable performance drop was observed when using YOLO to detect ECG sequences instead of referring to ground truth bounding boxes. In the 3x4 format with three rhythm signals (V1, II, and V5), YOLO frequently misses rhythm-V1 and rhythm-II signals. For noisy images, YOLO was able to detect the signals but often assigned incorrect tags. In this case, a perfectly digitized signal could receive a poor SNR score if it was assigned to the wrong ECG lead.

In Table 4, our classifier performance on the hidden validation is better than on our internal evaluation. Table 5 exhibits multiple performance differences across labels and training datasets. Our model significantly underperformed in CODE15%, which is the noisier dataset and less diverse in labels. Besides, some labels, such as Acute MI or PAC, are more difficult to classify. Their low numbers of positive samples play a role in this underperformance, which could be aggravated by complex ECG patterns such as in Acute MI.

4. Discussion and Conclusions

In this study, we proposed a dual deep-learning approach to digitize and classify 12-lead ECG images. Our digitization system comprised a fine-tuned YOLOv7 model to detect the ECG signals and a CNN-based model with self-attention to digitize the 10s signals. Our classi-

Training	Validation	Test	Ranking
0.728	0.742	-	5/90

Table 4. Macro F-measure for our selected entry (team Inventec AIC) on the classification task, including the ranking of our team on the hidden test. We show the score on our training hold-out fold, the best scoring on the hidden validation set, and one-time scoring on the hidden test.

	CODE15%	PhysioNet 2021	PTB-XL
AFIB/AFL (10.2%)	0.871	0.940	0.914
Acute MI (0.6%)	-	0.349	0.415
BRADY (12.9%)	0.508	0.967	0.606
CD (14.8%)	0.618	0.776	0.776
HYP (7.4%)	-	0.713	0.620
NORM (47.1%)	0.730	0.820	0.869
Old MI (5.4%)	-	0.623	0.740
PAC (2.0%)	-	0.678	0.654
PVC (2.6%)	-	0.718	0.851
STTC (13.3%)	-	0.672	0.739
TACHY (8.3%)	0.820	0.914	0.819
Macro Avg.	0.709	0.743	0.728

Table 5. F-measure per label on our internal evaluation. We included the percentage of positive samples of each label. The unlabelled classes of CODE15% were excluded.

fication model consisted of an EfficientNet-B0 trained for a multi-label task with sample weighting to handle class imbalance. Both models were pre-trained with numerous synthetic ECG images generated with more diverse ECG formats thanks to our extension of the generation code.

For the two-stage digitization, missed detections and inaccurate digitization have negative impacts on SNR, these combined errors could significantly lower the final score. On the other hand, since the fine-tuned YOLO was suspected to be biased in a specific format, including more data with various formats could improve the performance.

For the classification task, the additional training datasets with diverse ECG images considerably improved our performance. Our aggregation of predictions for long records and the probability thresholds were also relevant. However, our primary pursuit was to close the performance gap between classes. Our weighting strategies helped towards this objective, but our classifier was still underperforming with Acute MI and PAC. A possible improvement could be considering additional data points or patient information. Age and weight as risk factors of Acute MI could help detect more complex ECG patterns.

Another limitation of our approach is the lack of interaction between the digitization and classification models. A multi-task model could lead to beneficial relations between tasks. For example, the classification task would highlight the presence of ECG arrhythmic patterns on which digitization should focus.

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] 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.

[4] Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning. PMLR, 2019; 6105–6114.

[5] 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.

[6] 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`.

[7] 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;7:154.

[8] Strodthoff N, Mehari T, Nagel C, Aston PJ, Sundar A, Graff C, et al. PTB-XL+, a comprehensive electrocardiographic feature dataset. Scientific Data 2023;10:279.

[9] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[10] Reyna MA, Sadr N, Alday EAP, Gu A, Shah AJ, Robichaux C, et al. Issues in the automated classification of multilead ECGs using heterogeneous labels and populations. Physiological Measurement 2022;43(8).

[11] Ribeiro AH, Paixao G, Lima EM, Ribeiro MH, Pinto Filho MM, Gomes PR, et al. CODE-15%: A large scale annotated dataset of 12-lead ECGs (1.0.0), 2021. Zenodo.

[12] Cuturi M, Blondel M. Soft-dtw: a differentiable loss function for time-series. In International conference on machine learning. PMLR, 2017; 894–903.

Address for correspondence:

Chun-Ti Chou

111 No. 166, Sec. 4, Chengde Rd., Shilin Dist., Taipei, Taiwan

chou.peter@inventec.com