ECGFusion: Adaptive Multi-Model Framework for Noise-Robust Digitization and Fine-Grained Classification of Paper ECGs

Runze Shen[1,2], Ke Jiang[1,2], Sibo Wang[1,2], Peng Wang[1], Zhenfeng Li[1], Pang Wu[1], Xianxiang Chen[1], Lidong Du[1], Zhen Fang[1,2,3]

[1]State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

[2]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences; Beijing, China

[3]Research Unit of Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, Beijing, China

Abstract

This study proposes a multi-model fusion approach for digitization and classification of paper electrocardiograms (ECGs). We employ You Only Look Once version 10 (YOLOv10) for ECG signal area detection, DPLinkNet34 for image binarization, and GoogLeNet for multi-label classification. Our method excels in processing ECG images of varying quality, particularly those with noise and distortions. In the official phase of the TimeBeater team, on the hidden test data, we obtained a signal-to-noise ratio (SNR) of -0.056 for the digitization task, ranking 8th out of 16 teams, and a macro F-measure of 0.183 for the classification task, ranking 13th out of 16 teams. These scores are consistent with the published scores on the Challenge website. Additionally, on 5-fold cross-validation of the training set, we achieved an SNR of 7.273 and a macro F-measure of 0.833, though these cross-validation results are not comparable to the official challenge scores. This study is part of the "Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024".

1. Introduction

Electrocardiogram (ECG) analysis plays a crucial role in cardiovascular disease diagnosis, but the digitization and classification of paper ECG records remain challenging. This study presents ECGFusion, an innovative deep learning framework for digitization and classification of paper ECGs. Our approach integrates advanced computer vision and signal processing techniques, utilizing tools and technologies provided by ECG-Image-Kit[1] and leveraging resources from PhysioNet[2–4].

For the digitization task, we designed an ECG signal area detection model based on YOLOv10[5], introducing a dynamic threshold selection mechanism to improve signal localization accuracy in complex backgrounds. Additionally, we developed an improved DPLinkNet34 model[6] for ECG image segmentation and binarization, and implemented a weighted average path algorithm to effectively handle noise and signal interruptions. For the classification task, we proposed a multi-label classification model based on GoogLeNet[7], combining MultiLabelBinarizer and adaptive loss functions to significantly enhance the accuracy of ECG multi-pathology state classification. Notably, we designed an innovative HWDownsampling layer based on Haar wavelet transform[8] for image downsampling, effectively preserving key features of ECG signals. By implementing an innovative multi-model fusion strategy, we successfully integrated these components, achieving end-to-end processing from ECG images to multi-lead signal reconstruction and classification. These innovations fully utilize the advantages of deep learning in image processing and classification tasks, providing a comprehensive and efficient solution for ECG analysis.

2. Methods

This chapter provides a detailed description of our proposed method for intelligent digitization and classification of paper-based ECGs. Our system comprises two main modules: an ECG digitization module and an ECG image classification module. The digitization module, being the core of the system, incorporates multiple sub-models and complex processing workflows, while the classification module focuses on multi-label image classification tasks.

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2024.167

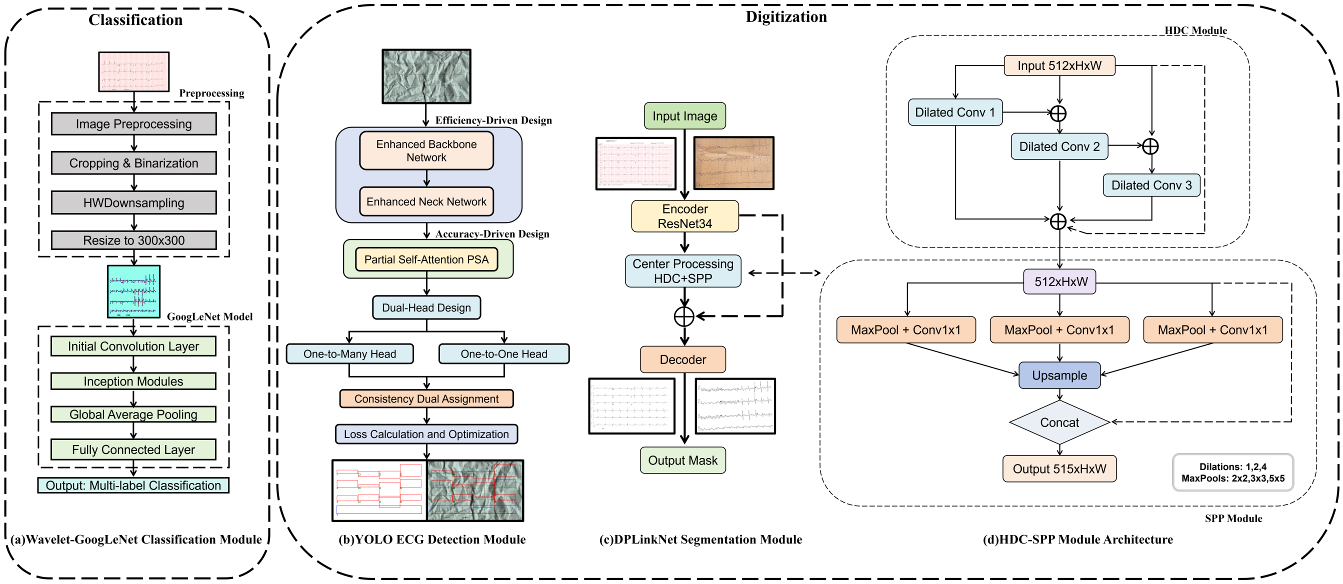

Figure 1. Overview of Key Architectures in Our ECG Classification and Digitization Framework. **(a)** Classification module with Wavelet-GoogLeNet. **(b)** YOLO ECG Detection module. **(c)** DPLinkNet Segmentation module. **(d)** Detailed view of the HDC-SPP Module Architecture within DPLinkNet. This framework integrates advanced techniques for comprehensive ECG analysis and processing.

2.1. Dataset and Pre-processing

We utilized the ECG-Image-Kit[1] to generate a diverse dataset of 50,000 ECG images, including clean, noise-induced, and artifact-induced variations. The dataset was annotated for multiple tasks: YOLOv10 (bounding boxes for ECG signal regions and lead labels), Lead Recognition (individual lead images with corresponding identifiers), DPLinkNet34 (clean, binarized images for segmentation ground truth), and Classification (multi-label annotations for cardiac conditions). Pre-processing involved image standardization, noise reduction, contrast enhancement, and data augmentation techniques.

2.2. System Architecture

Our system adopts a modular design, consisting of two core components: a multi-model fusion architecture for ECG signal digitization and a deep learning model for ECG image classification. The digitization module integrates object detection, lead recognition, image segmentation, and signal processing techniques, while the classification module concentrates on multi-label image classification tasks. This design allows for independent optimization of both modules while enabling them to work collaboratively within the overall system. Figure 1 illustrates our system's overall architecture, including key steps in the digitization module and the structure of the classification module. Specifically, Figure 1(a) shows the Wavelet-GoogLeNet architecture of the classification module, while Figures 1(b)-(d) detail the YOLO ECG detection, DPLinkNet segmentation, and HDC-SPP module architectures within the digitization module.

2.3. ECG Digitization

The ECG digitization module employs a multi-stage processing pipeline, combining computer vision and signal processing techniques. This module incorporates YOLOv10 for ECG signal region localization[5], a convolutional neural network for lead recognition, and DPLinkNet34 for image segmentation and binarization[6]. We preprocess the input image, including skew correction using a Fourier transform-based method[9] and adaptive compression to standardize the image size.

As shown in Figure 1(b), we employ an improved YOLOv10 model for ECG signal region localization. This model implements a dynamic threshold selection mechanism, optimizing detection results by iterating through different confidence thresholds. Concurrently, we integrate a specialized lead recognition network, a critical step in the digitization process. This network is designed with two convolutional layers, two max-pooling layers, a fully connected layer, and finally, an output layer with 12 neurons corresponding to 12 lead labels. Trained on a large corpus of synthetic data, this model not only enhances accuracy and robustness in recognizing various lead labels but also confirms the specific lead of the region localized by

YOLO detection, providing crucial information for subsequent signal processing.

The detected ECG regions are subsequently processed by the DPLinkNet34 model for segmentation and binarization, as illustrated in Figure 1(c). The DPLinkNet34 model combines depth-wise separable convolutions and a Spatial Pyramid Pooling (SPP) module, effectively enhancing feature extraction capabilities and receptive field. Figure 1(d) showcases the model's core components: the HDC (Hierarchical Dilated Convolution) module and the SPP module, which jointly enhance the model's ability to process multi-scale features[6]. The forward propagation of the HDC module can be represented as:

$$\text{HDC}(x) = x + D_1(x) + D_2(D_1(x)) + D_3(D_2(D_1(x))) \tag{1}$$

where D_1, D_2, and D_3 represent convolution operations with different dilation rates. In the signal extraction phase, we developed a weighted average path-based algorithm to effectively handle noise and signal interruptions. The extracted signals undergo smoothing using a Savitzky-Golay filter and voltage calibration based on pre-computed unit information, ensuring signal precision.

2.4. ECG Image Classification

The ECG classification module uses an improved GoogLeNet architecture[7] for multi-label image classification tasks. The preprocessing pipeline includes image preprocessing, cropping, binarization, and resizing. We introduce a Haar wavelet transform-based HWDownsampling layer for efficient downsampling while preserving key features[8]. HWDownsampling can be expressed as:

$$\text{HWDownsampling}(x) = \text{Conv}(\text{DWT}(x)) \tag{2}$$

The GoogLeNet model architecture includes initial convolution layers, Inception modules, global average pooling, and fully connected layers. The model's output layer is designed to accommodate multi-label output, with its dimensions matching the predefined number of ECG categories. We utilize MultiLabelBinarizer for label encoding, effectively handling the common occurrence of multiple pathological states in ECGs.

2.5. Model Training

We designed unique training strategies for the digitization and classification modules to address the characteristics and challenges of their respective tasks. The training of the digitization module is divided into multiple phases, optimizing YOLOv10, lead recognition network, and DPLinkNet34 separately.

YOLOv10 employs a novel consistent dual assignments strategy, combining one-to-many and one-to-one label assignment methods. During training, the one-to-many branch provides rich supervisory signals, while the one-to-one branch is used for efficient end-to-end inference. This approach maintains competitive performance with low inference latency.

The DPLinkNet34 model uses dice_bce_loss as its loss function, which is a combination of Dice loss and binary cross-entropy loss[10]:

$$L_{dice\_bce} = \alpha L_{dice} + (1 - \alpha) L_{bce} \tag{3}$$

The lead recognition network utilizes a classification cross-entropy loss function and is trained on a large corpus of synthetic data to ensure accuracy in real-world applications. The classification cross-entropy loss function for the lead recognition network can be expressed as:

$$L = -\sum_{i=1}^{12} y_i \log(\hat{y}_i) \tag{4}$$

where y_i is the true label and $\hat{y}_i$ is the model's predicted probability.

The classification module adopts a cross-validation strategy, utilizing the AdamW optimizer[11] with an initial learning rate of 0.001 and weight decay of 1e-5. We implement a cosine annealing strategy for dynamic learning rate adjustment:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + \cos(\frac{T_{cur}}{T_{max}}\pi)) \tag{5}$$

where η_t is the learning rate at time t, η_{min} and η_{max} are the minimum and maximum learning rates, T_{cur} is the current epoch, and T_{max} is the total number of epochs.

Through this meticulously designed training process, our models are capable of addressing complex ECG image digitization and classification tasks.

3. Results

Our method was evaluated on the hidden test set and the public training set. Table 1 summarizes our Challenge scores for the digitization and classification tasks on the hidden test set.

Task	Score	Rank
Digitization	SNR: -0.056	8/16
Classification	F-measure: 0.183	13/16

Table 1. Signal-to-noise ratio (SNR) and macro F-measure of our team's (TimeBeater) model on the hidden test data for the digitization and classification tasks, respectively.

These scores are consistent with the published scores on the Challenge website.Additionally, we performed 5-fold cross-validation on the public training set. The results are presented in Table 2.

Task	Cross-validation Score
Digitization	SNR: 7.273
Classification	F-measure: 0.833

Table 2. 5-fold cross-validation results on the public training set for our team's (TimeBeater) model.

It is important to note that these cross-validation results are not comparable to the official challenge scores and are provided only for additional context and insight into our method's performance on the training data.

4. Discussion and Conclusions

This study introduced ECGFusion, a multi-model approach for paper ECG digitization and classification. Our results highlight both the method's potential and limitations in real-world applications.

We achieved strong performance on the public training set using 5-fold cross-validation (SNR: 7.273 for digitization, F-measure: 0.833 for classification). However, performance dropped significantly on the hidden test set (SNR: -0.056, rank 8/16; F-measure: 0.183, rank 13/16), indicating challenges in generalization.

Our method's strengths include its modular design and innovative components like dynamic threshold selection in YOLOv10 and the HWDownsampling layer. Despite strong training set performance, the test set results reveal potential overfitting and computational complexity issues.

Future work should focus on expanding the dataset, implementing advanced augmentation techniques, exploring transfer learning, and developing lightweight architectures. These improvements could enhance ECGFusion's clinical utility, potentially improving cardiovascular disease diagnosis accuracy and efficiency.

Acknowledgments

This study was supported by the National Natural Science Foundation of China [grant number 62331025, 62071451,U21A20447], the National Key Research and Development Project (2021YFC3002204), and CAMS Innovation Fund for Medical Sciences [grant number 2019-I2M-5-019].

References

[1] Shivashankara KK, Shervedani AM, Clifford GD, Reyna MA, Sameni R, et al. ECG-Image-Kit: A Synthetic Image Generation Toolbox to Facilitate Deep Learning-Based Electrocardiogram Digitization. Physiological Measurement 202445(5):055019.

[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 2000101(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] 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`.

[5] Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, Ding G. YOLOv10: Real-Time End-to-End Object Detection. arXiv preprint arXiv240514458 2024.

[6] Xiong W, Jia X, Yang D, Ai M, Li L, Wang S. DP-LinkNet: A Convolutional Network for Historical Document Image Binarization. KSII Transactions on Internet and Information Systems TIIS 202115(5):1778–1797.

[7] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going Deeper With Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015 1–9.

[8] Li J, Cheng B, Chen Y, Gao G, Shi J, Zeng T. EWT: Efficient Wavelet-Transformer for Single Image Denoising. Neural Networks 2024177:106378.

[9] Pham L, Hoang PH, Mai XT, Tran TA. Adaptive Radial Projection on Fourier Magnitude Spectrum for Document Image Skew Estimation. In 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022 1061–1065.

[10] Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016 565–571.

[11] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. arXiv preprint arXiv171105101 2017.

Address for correspondence:

Zhen Fang

State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, No.19, North Fourth Ring West Road, Zhongguancun, Haidian District, Beijing, China

zfang@mail.ie.ac.cn