Long-term ECG Analysis Through Image Conversion and Deep Learning

Carlos Hernández-Fernández, Hilario Gómez-Moreno, Roberto Holgado-Cuadrado, Manuel Blanco-Velasco

Universidad de Alcalá, Alcalá de Henares, Spain

Abstract

Due to the lengthy information, irregular structure, and numerous types of distortion and disconnections, analyzing long ECG signals can be difficult. An ECG-to-image conversion is built enabling the use of image processing methods and tools for semantic segmentation that uses neural networks. That segmentation is based on the ResNet50 network, which is incorporated into the DeepLabv3 architecture, and is used to identify ventricular beat fusions and premature ventricular contractions. An established database is used for training and to assess the model's efficacy. The ECG data are converted into images using thresholding and heat maps, and PyTorch is used to train the network. The obtained binary masks mark the found abnormalities with accuracy. The presented method can help to detect heart anomalies more quickly and accurately by automating the study of long-term ECG readings.

1. Introduction

The semi-automatic processing of long ECG signals is a problem that has not yet been solved, due to the enormous amount of information to be processed, lacking uniform structure, with large doses of distortion of a diverse nature and periods of disconnection.

While cardiac abnormalities exhibit high variability among different ECGs or individuals, the proposed method enables the neural network to extract the key characteristics of the anomaly, thereby achieving generalization of its appearance given only a few examples.

The conversion of ECG into images has been achieved in the past, for example, with the use of spectrograms [1]. However, there are few examples where the conversion to images is done directly. One of them is Electrocardiomatrix (ECM) [2] that demonstrated improvements in the visualization and inspection of cardiac signals. This technique transforms a 2-dimensional ECG into a 3-dimensional matrix, enabling comprehensive analysis of ECG signals.

The presented work aims to establish an ECG-to-image conversion, similar to ECM but without the need for any synchronization, that enables the use of well established and proven image processing techniques and semantic segmentation tools to facilitate the detection of cardiac anomalies. Specifically, the focus lies on two specific abnormalities: premature ventricular contractions and ventricular beat fusions.

Premature ventricular complexes (PVCs) are common ventricular arrhythmias observed on ECGs [3]. While incidental PVCs are generally benign, frequent PVCs can negatively impact left ventricular function and contribute to heart failure. Also, PVCs can serve as a trigger for idiopathic ventricular fibrillation. A fusion beat is characterized by the coincidence of a supraventricular and a ventricular impulse, resulting in a hybrid complex. In the case of a ventricular fusion beat [4] or fusion PVC, this occurs when these electrical impulses coincide within the ventricular chambers. Fusion beats can be observed in conditions such as ventricular tachycardia and accelerated idioventricular rhythm.

2. Methodology

2.1. ECG to Image Conversion

The process starts applying upper and lower thresholds to prevent the loss of information near the signal's mean value. In this case, a 25% threshold has been applied. It is common to find peaks in ECGs (maxima and/or minima) of much larger dimensions compared to the true maxima and minima of the QRS complex, which are usually caused by the presence of noise. The presence of these disproportionate peaks in the subsequent conversion of the ECG to an image results in the loss of most of the genuine signal information.

In our case, each record was divided into segments of 10^6 samples and rearranged in a 1000×1000 matrix. Subsequently, each matrix was converted into a full image with dimensions of 1000×1000 pixels using a heat map that assign different colors to different ECG values. These images were then re-scaled to 2560×2560 pixels. The whole process is shown in Figure 1.

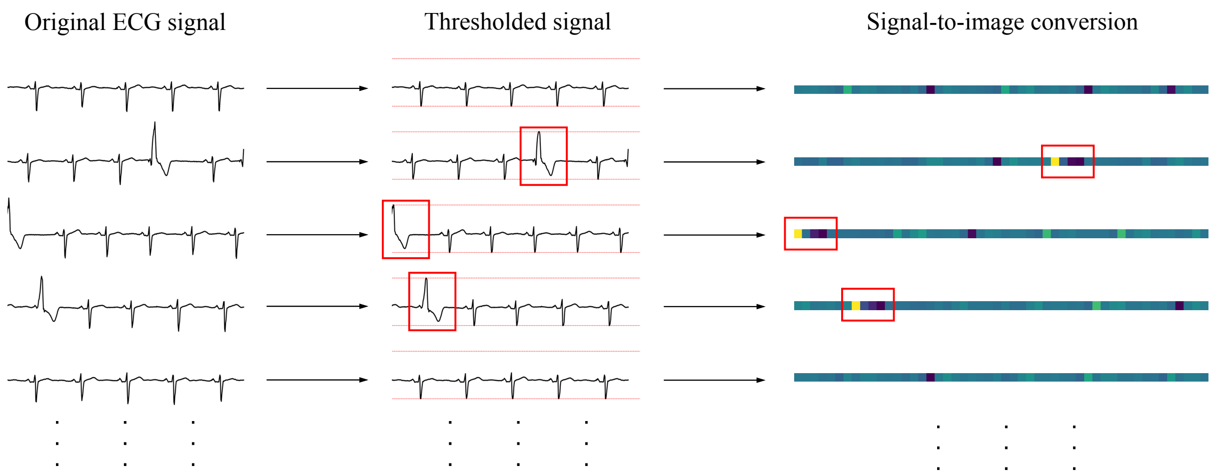

Figure 1. Conversion from ECG to image.

Then, using the provided labels by Physionet for this dataset, a binary segmentation mask was created from each image. This mask assigns a value of "1" (White) to the pixels corresponding to the cardiac anomalies targeted for detection. Conversely, pixels in the images that do not correspond to any anomaly are assigned a value of "0" (Black).

2.2. Network Specifications.

The ResNet50 network, a Residual Network with 50 layers, has emerged as one of the leading networks for semantic segmentation [5]. This work utilizes ResNet50 as the backbone of a DeepLabv3 architecture to process the ECG images. For the implementation of the network, the PyTorch library in Python was utilized.

For each of the training records used, a 2560×2560 image was created using their first 10^6 samples. Subsequently, a predefined number of patches was extracted from each of these images to build the training dataset. The test dataset was built using one of the training ECGs but starting from the 10^6th sample, thus excluding the data already used in the training dataset. Since the network expects input images of size 256×256 pixels [6], each image (and its corresponding mask) in the created dataset was segmented into 100 patches of size 256×256. Then the first 10, 20, 50 and 100 patches of size 256×256 were extracted, resulting in a total of 50, 100, 250 and 500 images that formed four different training datasets.

The implementation of the network utilizes the DeepLabv3 architecture [7], which is based on the concept of dilated convolutions and incorporates techniques such as Atrous Spatial Pyramid Pooling (ASPP). ResNet50 is chosen as the backbone of the DeepLabv3 architecture.

One of the key steps to ensure good convergence and learning of the model is the selection of the network's hyperparameters. The most important were the *learning rate* that influences the convergence and learning speed of the model, the *batch size* that represents the number of training examples used in each iteration and affects the trade-off between computation efficiency and model generalization and the *number of steps* that determines the total number of iterations during the training process. After conducting the necessary experiments, it has been determined that the optimal combination of hyperparameters consists of a learning rate of 10^{-4}, a batch size of 32 and 3000 training steps.

Another crucial parameter for the network is the optimizer, used to update the network's weights based on the computed gradients during backpropagation. In our work, the Adam optimizer [8] was used. This optimization algorithm that combines the benefits of the RMSprop and Momentum algorithms to improve the learning process of a model. It adapts the learning rate based on the distribution of parameters within the model.

Additionally, it is important to establish an appropriate loss function. The cross-entropy [9] is commonly used as a loss function in classification problems where the goal is to minimize the distance between the predicted probability distribution by the model and the actual distribution and has been used in this work.

Lastly, during the training, the Intersection over Union (IoU) metric is calculated. IoU metric is a measure of similarity between two sets of pixels in an image [10] and in the context of semantic segmentation networks is used to evaluate the quality of the segmentation obtained by the model.

Finally, the code used in this research is an adaptation of the one published in [11].

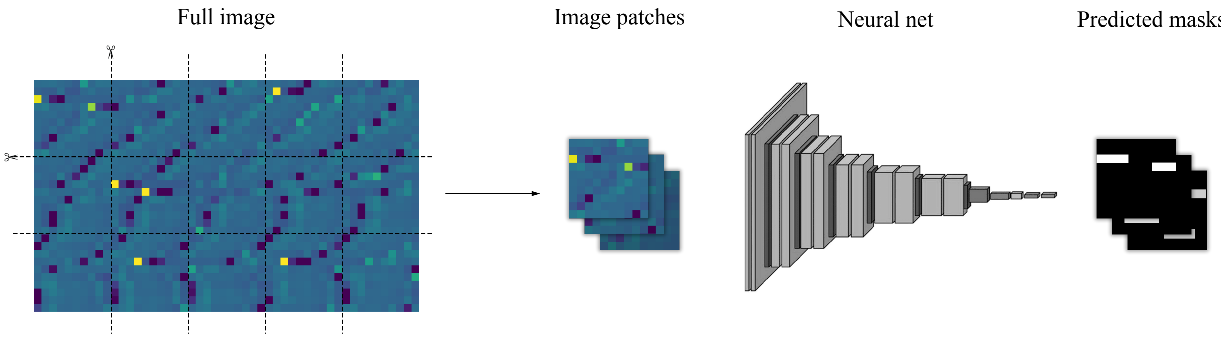

Figure 2. Segmentation process.

2.3. Experiment Database

For this research, the "MIT-BIH Long-Term ECG Database" [12] from Physionet has been employed. This database comprises 7 long-term ECG, each ranging from 14 to 22 hours, with manually revised beat annotations. Each of these long-term ECG recordings has been acquired at a sampling frequency of 128 samples per second. Only 5 of those 7 records were finally used since one of them was extremely noisy and the other one had no enough labeled anomalies.

2.4. Predicted Masks Generation

After the training, we have a model that processes 256×256 pixel input images and gets a predicted binary mask. Therefore, to process a complete ECG image, it is necessary to process each of its patches and then reassemble them into a complete image of 2560×2560 pixels. Figure 2 shows this process.

Once the predicted mask has been reconstructed, we can get the correspondence between pixels and the initial samples. To accomplish this, a tool has been implemented to convert pixel coordinates (height, width) to sample numbers and to extract the samples corresponding to anomalous detections. These samples are obtained as a range rather than a single sample, which is the way the signals are labeled in Physionet.

3. Experimental Results

The training phase of the network used an NVIDIA RTX 4080 graphics card with 16GB of memory. Using this card, the average convergence time for the neural network was approximately 20 minutes, while the average time for image processing and mask reconstruction was around 15 seconds. On the other hand, once the network has been trained, the processing of an ECG is not computationally expensive, so since there is no foreseen need for real-time implementation of the model, it would be feasible to employ only CPU for this purpose.

Once the network training is completed, after 3000 training steps, a mean IoU of 0.81 is achieved, while the cross-entropy loss function reaches a mean value of 0.05.

To assess the effectiveness of the network, one can compare the ground truth mask with the mask predicted by the network using image processing techniques and libraries.

The following experiments[1] have been conducted using record 14184 as the test set, excluding some initial samples. This initial sample set was gradually included in the training dataset to observe the effect it has on anomaly detection accuracy.

After the segmentation, a refinement of the masks predicted by the network was performed. This treatment adjusts the areas detected as anomalies by the network to well-defined rectangles similar to those present in the test masks. To evaluate the results, the four models (with different number of patches) mentioned earlier were employed. On one hand, pixel-level comparisons were made between the real masks (ground truth) and the masks predicted by the network.

Table 1. Detection results at pixel level.

Patches per ECG	Precision	Recall	F1
10	0.41344	0.60901	0.49252
20	0.40789	0.71681	0.51993
50	0.44037	0.77708	0.56216
100	0.45721	0.83405	0.59064

The obtained metrics (Table 1) range from 0.413 to 0.457 for Precision, from 0.609 to 0.834 for Recall, and from 0.492 to 0.590 for the F1 score. Pixel-level comparisons are too strict since the network is unable to provide pixel-perfect precision. For this reason, an analysis

[1] `https://github.com/hgomezmoreno/IMCG_CinC2023`

at the anomaly level was carried out. As shown in Table 2, the results at this level exhibit a high level of accuracy in anomaly detection, ranging from 76.5%, when 1.1% of each ECG is included in the training dataset (13 minutes), to 92.7% when 11.1% of each ECG is included (130 minutes).

Table 2. Detection accuracy at anomaly level. AT per ECG represents the average time per record in % included in the training dataset.

Patches per ECG	Time per ECG (min)	AT per ECG (%)	Detection (%)
10	13	1.1	76.5
20	26	2.2	83.7
50	65	5.5	88.1
100	130	11.1	92.7

4. Conclusions and Future Work

This study analyzes the possibility of employing image processing techniques and neural networks to detect cardiac anomalies.

While the pixel-level results deviate from an optimal outcome, when studying anomalies at a higher level, it is observed that the proposed network successfully detects a high percentage (between 76.5% and 92.7%) of anomalies when trained with a short segment of the ECG to be processed.

By leveraging the capabilities of the developed model, medical practitioners could save time and effort in ECG analysis while improving the overall efficiency and accuracy of the diagnostic process.

The main limitation of this study is the requirement to include an annotated segment of the ECG that needs to be processed. As shown in the results, the proposed model still exhibits some dependency on the input data. The short-term objective is to enhance the model to become fully agnostic to the specific ECG being processed, thereby eliminating the need to incorporate portions of that ECG into the training dataset.

Another purpose is to endow the model with a more diverse dataset, encompassing a larger number of patients with labeled anomalies. This, in conjunction with a multi-class semantic segmentation model, would enable the detection of multiple anomalies within a single model.

This work could potentially facilitate the implementation of a system that aids medical personnel in the automatic or semi-automatic analysis of long-term ECG signals. It could also assist medical professionals by identifying periods of time with a higher likelihood of detecting anomalies, thus reducing the challenge of processing long-duration ECGs to interpreting shorter-duration ECGs.

Acknowledgments

This work was supported by Universidad de Alcalá through the project PIUAH22/IA-023

References

[1] Huang J, Chen B, Yao B, He W. Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network. IEEE access 2019;7:92871–92880.

[2] Li D, Tian F, Rengifo S, Xu G, Wang MM, Borjigin J. Electrocardiomatrix: A new method for beat-by-beat visualization and inspection of cardiac signals. J Integr Cardiol 2015; 1(5):124–128. URL `https://doi.org/10.15761/JIC.1000133`.

[3] Hoogendijk MG, Géczy T, Yap SC, Szili-Torok T. Pathophysiological mechanisms of premature ventricular complexes. Frontiers in Physiology 2020;11. ISSN 1664-042X. URL `https://www.frontiersin.org/articles/10.3389/fphys.2020.00406`.

[4] Marriott HJ, Schwartz NL, Bix HH. Ventricular fusion beats. Circulation 1962;26(5):880–884.

[5] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 770–778.

[6] Pytorch Foundation. Resnet50 - torchvision.models. `https://pytorch.org/vision/main/models/generated/torchvision.models.resnet50.html`, 2023. Accessed on July 6, 2023.

[7] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation, 2017.

[8] Kingma DP, Ba J. Adam: A method for stochastic optimization, 2017.

[9] Mao A, Mohri M, Zhong Y. Cross-entropy loss functions: Theoretical analysis and applications, 2023.

[10] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015; 3431–3440.

[11] Eppel S. Train a neural net for semantic segmentation in 50 lines of code, with pytorch. Towards Data Science 2021; URL `https://shorturl.at/cfuDM`. Accessed on march 26, 2023.

[12] 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):215–220.

Address for correspondence:

Hilario Gómez-Moreno
Universidad de Alcalá, Departamento de Teoría de la Señal y Comunicaciones, 28871, Alcalá de Henares, Madrid, España
hilario.gomez@uah.es