U-Net Architecture for the Automatic Detection and Delineation of the Electrocardiogram

Guillermo Jimenez-Perez[1], Alejandro Alcaine[1], Oscar Camara[1]

[1]BCN MedTech Unit, PhySense Group, Department of Information and Communication
Technologies, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Automatic detection and delineation of the electrocardiogram (ECG) is usually the first step for many feature extraction tasks. Although deep learning (DL) approaches have been proposed in the literature, those employ non-optimal and outdated architectures. Thus, rule-based algorithms remain as state-of-the-art. Nevertheless, those may not generalize on other datasets and require difficult offline tuning for adapting to new scenarios. This work frames this task as a segmentation problem for using an adaptation of the U-Net architecture, a fully convolutional network. The detection performance shows a precision of 89.27%, 98.18% and 93.60% for the detection of the P, QRS and T waves, respectively, and a recall of 89.07%, 99.47% and 95.21%. This work shows promising results, outperforming existing DL approaches while being more generalizable than rule-based methods.

1. Introduction

Surface electrocardiogram (ECG) is currently the main tool for monitoring cardiac events and diagnosis of cardiac illnesses due to its widespread accessibility and easy usage. Some illnesses and cardiac abnormalities are subtle to the human eye and dependent of cardiologist's expertise. In this sense, data-based analysis of the ECG can provide more robust and objective insights of the underlying cardiac condition by extracting data-based clinical markers [1]. Moreover, automatization with machine learning might simplify tasks not requiring interpretation.

In this work, the automatic detection and delineation of the P, QRS and T waves in the ECG, a key step for further feature extraction, is explored. [1]. The most accurate algorithms available in the literature are based on the wavelet transform with rule-based adaptive thresholds for computing the fiducials [2–4]. These methods, however, may over-adjust to the whole dataset, thus compromising their generalization and jeopardising their performance on unseen examples. Alternatively, machine learning techniques have been proposed [5–7]. However, these methods have been recently surpassed by deep learning (DL) techniques, using convolutional neural networks (CNN) [8,9] and long short-term memory [10]. Nonetheless, these methods employ non-optimal and outdated architectures for this task, especially given the state of the art results using modern pattern recognition architectures for semantic segmentation or image classification.

In this work we present the adaptation of a U-Net architecture [11] to 1D data aiming at detecting and delineating the P, QRS and T waves of the ECG.

2. Materials and Methods

2.1. QT database

The QT database [12] has been employed as ground truth for training and evaluating the model. This database consists of 105 ambulatory two-lead recordings of 15 minutes of length, with a sampling frequency of 250 Hz. Considering both leads, the database contains a total of 210 single-lead ECGs. For every lead, the database holds nine different fiducials: the P, QRS and T wave markers for detection and their respective onset and offset landmarks for delineation.

2.2. Database split

The database was split intro training, testing and validation sets. Additionally, the analysis of ECG recordings usually requires windowing the signal into smaller chunks. For ensuring that the sets are non-overlapping database splitting must be performed strictly before windowing to prevent the existence of different instances of the same recording (in either lead) in both training and testing sets, also called *subject-wise* splitting [13]. In this work, a "subject" is every element of the 105 recordings, comprising both leads, before windowing. A window and a stride of 2^{11} and 2^{10} samples were used, allowing for some overlap. This helps to learn the window's boundaries, thus alleviating the drawbacks of the windowing process.

ISSN: 2325-887X DOI: 10.22489/CinC.2019.284

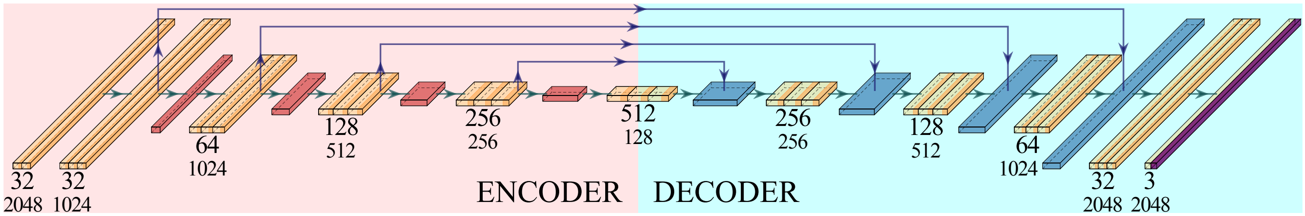

Figure 1. Proposed U-Net architecture. Orange blocks indicate stacked convolutional operations with ReLU activations. Red blocks indicate pooling operations. Blue blocks indicate upsampling operations followed by concatenation of the skipped connections (vertical arrows). The purple block indicates a sigmoid activation function. Horizontal arrows indicate connections between operations. The block numbering indicates number of channels (top) and the data size (bottom). The signal, as a 1D vector, is used as input in the leftmost convolutional layer and the output is retrieved at the rightmost.

2.3. Network architecture

The U-Net architecture is composed of an encoder, a decoder and skipped connections between the encoder and the decoder, as illustrated in Figure 1. The encoder is composed of several levels of convolutional operations, which extract increasingly abstract representations of the input data; and downsampling blocks, which reduce input complexity. The decoder block is composed of several layers of convolutional blocks, upsampling blocks and concatenation blocks. In U-Net architectures, the number of filters is doubled on each downsampling block and halved on every upsampling block for capturing as much information as possible within the available computational resources [11]. We define a level as a conjunction of convolutions followed or preceded by an up/downsampling layer.

Our implementation of the U-Net substitutes the original 2D convolutions with 1D operations. Also, the original U-Net cropped the input due to computational constraints, which is no longer relevant in this work due to increased computational power.

Convolutional blocks perform a trainable convolutional operation followed by a non-linear activation function:

$$\mathbf{y}[i,k] = \sum_{j=-K_s/2}^{K_s/2} \mathbf{x}[i+j,k]\mathbf{w}[j], \tag{1}$$

where $\mathbf{y}$ is the output of the convolution, $\mathbf{x}$ is the output of the previous layer, $\mathbf{w}$ is the trainable convolutional filter, $k \in [0,K]$ is the current feature map, $i \in [0,N]$ is the current element of $\mathbf{x}$, and K_s is the kernel size.

After this operation is performed $\forall i,k$, a non-linear activation function H, specifically a rectified linear unit (ReLU), is applied on the output of the convolution:

$$H(\mathbf{y}) = \begin{cases} \mathbf{y} \text{ if } \mathbf{y} \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Thus, the objective of the training procedure is to learn the set of weights $\mathbf{w}$ for every convolutional layer so that the combination of operations attains the final results.

2.4. Implementation details

The U-Net was instantiated for a total of 5 levels in the encoder and the decoder, with three convolutional layers per level. The last convolutional layer on the decoder has three output channels and a sigmoid activation. The output channels encode the independent detection and delineation of the P, QRS and T waves, respectively. The convolutional blocks used batch normalization as regularizer. The starting number of trainable filters was 2^5 and doubled after every downsampling operation.

The training procedure was performed on an Nvidia TITAN Xp, with a batch size of 2^9. Adam was used as optimizer with default parameters [14]. Xavier normal was used as the kernel initializer [15]. The model was implemented in Keras[1]. 5-fold cross-validation was performed for testing the generalizability of the approach, using the aforementioned *subject-wise* approach [13]. 20% of the recordings of the training set were used for validation to guide the training process.

2.5. Training metrics and post-processing

The usual metrics for CNN-based semantic segmentation are based on coefficients of overlap between masks such as the Dice or Jaccard indices. For using these metrics, the ground truth was transformed from the P, QRS and T fiducials to one mask per wave, valued $True$ from the onset to the offset of the wave and $False$ otherwise.

The architecture, thus, outputs one dense binary mask per wave, as it can be seen in Figure 2 (top-right panel). These masks are transformed into its original fiducials (Figure 2, top-left panel) by identifying each dense trail as a wave for detection, and the samples where the trail begins and ends as its onset and offset for delineation. To avoid performance degradation due to noisy border definitions, a closing operation of 5 samples was applied to each predicted mask before their conversion to fiducials.

[1] https://keras.io/

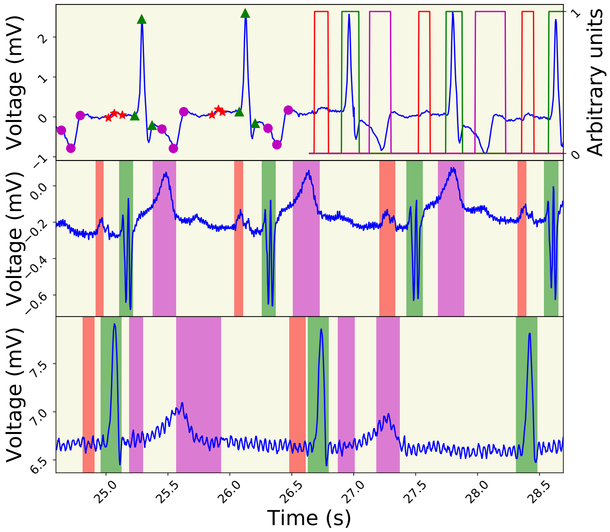

Figure 2. Detection example. Ground truth, as fiducials (top-left) and as mask (top-right). Successful prediction (middle) overlaid on the background. Unsuccessful prediction (bottom) overlaid on the background. Red: P wave. Green: QRS wave. Magenta: T wave.

2.6. Evaluation

The metrics for evaluation were the precision and recall of every wave separately as $P^+ = TP/(TP + FP)$ and $Re = TP/(TP + FN)$, where a true positive (TP) was counted whenever the wave fiducial in the ground truth was contained between the predicted onset and offset of that wave, and a false negative (FN) otherwise. A false positive (FP) is considered whenever a predicted wave fiducial was not contained between the onset and offset of the ground truth. For every TP, the mean and standard deviation (SD) of the error on the onset and offset location are calculated with respect to the ground truth.

3. Results and discussion

Table 1 shows the detection performance of this work compared against [3, 9]. The average precision and recall are 93.68% and 94.59%, respectively. The error in the onset and offset locations were shown in Table 2, compared against [3, 8]. Figure 2 shows detection examples.

The precision of the P and QRS waves are comparable to [3], whereas the T wave deviates slightly. The recall of the P and T waves is worse than [3]. This trend is maintained in the identification of the onsets and offsets. When compared to DL methods [9], our network outperforms or has comparable detection performance, except when using a tolerance window of 150ms. On the other hand, our delineation results are poorer than [8]. However, our network also provides delineations for the P and T waves.

Although this work does not outperform [3], DL methodologies have the potential advantage of adapting to new morphologies with minimal human intervention, given run-time corrections, as opposed to rule-based approaches, which require careful offline calibration.

This work contributes by framing the ECG detection and delineation as a segmentation problem, which has the advantage of leveraging the richness of mature architectures employed in that area; especially by alleviating the main image segmentation bottleneck, the number of parameters, due to the relative simplicity, parameter-wise, of 1D data. Moreover, DL can benefit from transfer learning, which enhances the outcomes of the models by pre-training them on data from another domain. Even though transfer learning is being widely applied in image analysis, it can be further exploited in 1D signal applications.

All these advantages allow the construction of more complex networks which might stretch a model's performance, potentially unburdening physicians of more mechanical work, such as ECG delineation an ECG or monitoring a patient's condition.

However, our work shows certain limitations. The onset and offset delineation shows a high SD error, especially in the T wave. We hypothesize that the performance of the T wave is hindered due to high input variability and the network's relative simplicity; and that the general performance is affected by cross-validation. Future work will enhance the architecture towards more reliable results. The possibility of adding residual connections [16] or separable convolutions [17] are natural extensions of this work.

4. Conclusions

This work shows a promising step towards an automatic delineation of the ECG by adapting a U-Net, a well-known

Table 1. Detection results for the P-QRS-T waves.

	Precision (%)					Recall (%)				
	This work	[3]	[9]			This work	[3]	[9]		
Wave			10 ms	50 ms	150 ms			10 ms	50 ms	150 ms
P	89.27	91.03	79.6	84.6	90.0	89.07	98.87	86.8	92.2	98.1
QRS	98.18	99.86	93.0	98.5	99.9	99.47	99.80	92.2	97.7	99.1
T	93.60	97.79	80.2	87.4	97.7	95.21	99.77	80.7	87.9	98.3

Table 2. Delineation error (mean±SD) of the onset and offset landmarks for the different waves. N/A stands for not applicable.

	Onset (ms)			Offset (ms)		
Wave	This work	[3]	[8]	This work	[3]	[8]
P	-3.3±15.8	2.0±14.8	N/A	6.8±15.2	1.9±12.8	N/A
QRS	-3.2±22.4	4.6±7.7	-2.6±10.8	7.9±21.9	0.8±8.7	4.4±15.2
T	8.3±52.6	0.2±13.9	N/A	7.4±59.8	-1.6±18.1	N/A

DL architecture designed for image analysis, to 1D data, offering nearly state-of-the-art results. The ability of this network to recognize patterns in the input data makes it a natural choice for working on the delineation of the ECG, task that does not rely heavily on temporal dependencies, for which other architectures such as LSTM are more adequate.

Acknowledgments

This research was supported by the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia and European Social Fund (2017 FI_B 01008). This work is part of ATTRACT (MERIT-VA Project) that has received funding from the European Union's Horizon 2020 Research and Innovation Programme. The TITAN Xp used for this work was donated by the NVIDIA Corporation.

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. J R Soc Interface 2018;15(138):20170821.

[2] Li C, Zheng C, Tai C. Detection of ECG characteristic points using wavelet transforms. IEEE T Biomed Eng 1995; 42(1):21–28.

[3] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. IEEE T Biomed Eng 2004;51(4):570–581.

[4] Alcaine A, Soto-Iglesias D, Calvo M, Guiu E, Andreu D, Fernandez-Armenta J, Berruezo A, Laguna P, Camara O, Martinez JP. A wavelet-based electrogram onset delineator for automatic ventricular activation mapping. IEEE T Biomed Eng 2014;61(12):2830–2839.

[5] Dubois R, Maison-Blanche P, Quenet B, Dreyfus G. Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators. Comput Meth Prog Bio 2007;88(3):217–233.

[6] Graja S, Boucher JM. Hidden Markov tree model applied to ECG delineation. IEEE T Instrum Meas 2005;54(6):2163–2168.

[7] de Lannoy G, Frenay B, Verleysen M, Delbeke J. Supervised ECG delineation using the wavelet transform and hidden Markov models. In IFMBE Proc. Springer, 2008; 22–25.

[8] Camps J, Rodriguez B, Mincholé A. Deep learning based QRS multilead delineator in electrocardiogram signals. In CinC, volume 45. IEEE, 2018; 1–4.

[9] Sodmann P, Vollmer M, Nath N, Kaderali L. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. Physiol Meas 2018; 39(10):104005.

[10] A. Hedayat, H. Chia, X. Zhou, C. Richard. CM. Supervised ECG interval segmentation using LSTM neural network. In BIOCOMP. ACSE, 2018; 71–77.

[11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In MICCAI, volume 9351. LNCS, 2015; 234–241.

[12] Laguna P, Mark R, Goldberg A, Moody G. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In CinC, volume 24. IEEE, 1997; 673–676.

[13] Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. Comput Meth Prog Bio 2018;161:1–13.

[14] Kingma DP, Ba J. Adam: a method for stochastic optimization. In ICLR. arXiv, 2015; 1–15.

[15] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In AISTATS, volume 9. JMLR, 2010; 249–256.

[16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In CVPR. IEEE, 2016; 770–778.

[17] Chollet F. XCeption: deep learning with depthwise separable convolutions. In CVPR. IEEE, 2017; 1800–1807.

Address for correspondence:

Guillermo Jimenez-Perez
Carrer Tànger 122-140, Office 55.119
08018 Barcelona, Spain
guillermo.jimenez@upf.edu